Hierarchical Bayesian Space-Time Models

Christopher K. Wikle*

Geophysical Statistics Project

National Center for Atmospheric Research

L. Mark Berliner

National Center for Atmospheric Research

and Ohio State University

Noel Cressie

Iowa State University

Abstract

Space-time data are ubiquitous in the environmental sciences. Often, as is the case with atmospheric and oceanographic processes, these data contain many different scales of spatial and temporal variability. Such data are often non-stationary in space and time and may involve many observation/prediction locations. These factors can limit the effectiveness of traditional space-time statistical models and methods. In this article, we propose the use of hierarchical space-time models to achieve more flexible models and methods for the analysis of environmental data distributed in space and time. The first stage of the hierarchical model specifies a measurement-error process for the observational data in terms of some "state" process. The second stage allows for site-specific time series models for this state variable. This stage includes large-scale (e.g., seasonal) variability plus a space-time dynamic process for the "anomalies". Much of our interest is with this anomaly process. In the third stage, the parameters of these time series models, which are distributed in space, are themselves given a joint distribution with spatial dependence (Markov random fields). The Bayesian formulation is completed in the last two stages by specifying priors on parameters. We implement the model in a Markov chain Monte Carlo framework and apply it to an atmospheric data set of monthly maximum temperature.

Keywords: atmospheric science, dynamical systems, environmental studies, Gibbs sampling, Markov random field, MCMC, non-stationarity, temperature

*Address: Box 3000, Boulder, Colorado 80307. (303) 497-1722, email: wikle@ucar.edu

1 Introduction

Many environmental processes involve variability over space and time. For example, meteorologists are typically interested in the time evolution of certain atmospheric variables (e.g., wind, temperature, moisture) over specified spatial domains. One need only examine the governing hydrodynamical equations of the atmosphere (e.g., Holton 1992) to see that such dynamical processes must involve complicated spatial structure, temporal structure, and spatio-temporal interactions. However, the statistical characterization of such complicated processes via traditional space-time approaches is hampered by the need to specify these space, time, and space-time interaction components of variation. The space-time variability is further complicated by the notion that there can often be very different spatial behavior at different points in time, as well as different temporal variability at different locations in space. These difficulties, plus the large data sets typical of many environmental problems, often require practitioners to apply potentially unrealistic simplifying assumptions. Although we cannot escape the "curse of dimensionality", we can take advantage of recent increases in computational speed and numerical advances (e.g., Markov chain Monte Carlo) that allow us to implement Bayesian space-time dynamical models in a hierarchical framework. Such specifications provide simple strategies for incorporating complicated space-time interactions at different stages of the model's hierarchy, and the models are feasible to implement in high dimensions.

There are several examples of hierarchical space-time models applied to environmental problems. For instance, Hughes and Guttorp (1994) used hidden Markov models with unobserved weather states to model space-time atmospheric precipitation. Brown et al. (1994) applied a hierarchical model to a relatively low dimensional space-time air pollution problem.

Similarly, several recent examples of Bayesian space-time modeling have been reported in the literature. Hainsworth and Mardia (1992) extended the Bayesian Markov random field (MRF) image restoration work of Geman and Geman (1984) to image sequences. Handcock and Wallis (1994) used a Bayesian kriging approach to space-time modeling of meteorological fields. In addition, Huang and Cressie (1996) and Wikle (1996) developed empirical Bayesian space-time Kalman filter models for the investigation of snow water equivalent and monthly precipitation,

respectively. Waller et al. (1997) employed hierarchical Bayesian space-time models for mapping disease rates.

In this article we illustrate the Bayesian hierarchical view in space-time settings. A flexible, five-stage hierarchical model is presented. We let Y, later indexed by both space and time, represent the primary variable of interest. The first stage of the model specifies a measurement error model for the observational data, denoted by Z; typically, Z is an observation of Y with error. The second stage of the model allows for (i) site-specific time series models for Y, and (ii) the incorporation of space-time dynamics. These dynamical terms are denoted by X. In the third stage, the parameters of the site-specific time series models are themselves endowed with priors (Markov random fields) that generate spatial dependence structures. Also, the dynamical terms X are modeled in the third stage. This aspect is central to this article and distinguishes our approach from other hierarchical space-time formulations (Gelfand et al. 1995; Waller et al. 1997). The final two stages complete the Bayesian formulation by specifying priors on parameters.

In some sense we re-pose the fundamental challenge of space-time modeling: How does one effectively model spatial and temporal structures? The Bayesian hierarchical strategy that we propose allows complicated structure to be modeled in terms of means at various stages, rather than a model for a massive joint covariance matrix. Furthermore, the models proposed here offer the analyst various opportunities to explore and employ trade-offs between rich time-lagged, though site-specific, time series models, and models based on more direct specifications of spatial structures, but at short time lags.

We believe that the class of models discussed here offer a useful framework for many settings. Nevertheless, the models are not the most general. A variety of additions or variants are possible. Issues we considered in developing this model include implementations in very large data problems and characterization of underlying space-time dynamic processes (see West and Harrison (1989) for an overview of Bayesian dynamic models). In particular, we envision space-time data sets involving many observation/prediction locations and potentially very long data records, so that models with space, time, and space-time interaction structure are needed. We also require models that are capable of handling a variety of features occurring at varying space and time scales. Third, relative ease in computation was considered. While Markov chain Monte Carlo (MCMC)

is a powerful and useful technique, rich models in very high dimensions lead to technical problems in its implementation. Hence, the very large data settings we consider require some concessions in model generality.

2 Notation

Suppose $Y(s,t)$ denotes the value of the process of interest at location s and time t, where $(s,t) \in \mathcal{M}$ and $\mathcal{M}$ is a lattice or grid in space-time. The resolution of $\mathcal{M}$ must be sufficiently rich to allow for desired predictions, interpolations, etc. An alternative view begins with models that are indexed by continuous space and time variables. One can then analyze the implied behaviors at the grid level.

Assume that there are S sites and T time points, and hence the number of grid points is $S \times T$. The notation s for sites can be refined to allow spatial information to be specified. For example, in two spatial dimensions we could write $s = (k,l)$, in Cartesian coordinates. However, in this section and the next, it is convenient to continue with the coordinate-free notation.

Depending on the modeling strategy, it may be helpful to "arrange" the process in different ways. To that end, let $\mathbf{Y}(s,)$ be the T-vector of time series values at site s. Let $\mathbf{Y}(,t)$ be the matrix of spatially gridded values of the process at time t. Let $\vec{\mathbf{Y}}_t$ denote a vectorization of the matrix $\mathbf{Y}(,t)$. Also, we let

$$Y = \{Y(s,t) : (s,t) \in \mathcal{M}\}.$$

Hierarchical modeling involves the specification of a series of conditional models. As we develop our formulation, we let θ, subscripted by the stage number at which it is first introduced as a condition, represent a collection of model parameters. Other symbols are reserved for primary variables that are the focus of the modeling strategy. Also, throughout the paper, we use the following notation, popularized by Gelfand and Smith (1990): For a random vector $\mathbf{U}$, let $[\mathbf{U}]$ represent the joint probability density function of $\mathbf{U}$, and let $[\mathbf{U}|\mathbf{V}]$ represent the conditional density of $\mathbf{U}$, given the vector $\mathbf{V}$.

3 A Hierarchical Space-Time Model

A casual summary of the five stages of the basic hierarchical space-time model is presented in Table 1 for convenience.

Table 1: Hierarchical Model

Stage	Variables	Model	Sub-model
1	Data	$[Z\|Y,\theta_1]$	
2	Process	$[Y\|\mu,\beta,X,\theta_2]$	
3	Large & small scales	$[\mu,\beta,X\|\theta_3=(\theta_\mu,\theta_\beta,\theta_X)]$	
	Spatial prior: means		$[\mu\|\theta_\mu]$
	Spatial prior: seasonalities		$[\beta\|\theta_\beta]$
	Space-time dynamics		$[X\|\theta_X]$
4	Model parameters	$[\theta_1,\theta_2,\theta_3\|\theta_4=(\theta_4(1),\theta_4(2),\theta_4(3))]$	
	Measurement variances		$[\theta_1\|\theta_4(1)]$
	Model variances		$[\theta_2\|\theta_4(2)]$
			$[\theta_\mu\|\theta_4(\mu)]$
			$[\theta_\beta\|\theta_4(\beta)]$
	Dynamical parameters		$[\theta_X\|\theta_4(X)]$
5	Hyperparameters	$[\theta_4]=[\theta_4(1)][\theta_4(2)][\theta_4(\mu)][\theta_4(\beta)][\theta_4(X)]$	

3.1 First Stage: Measurement Process

Let Z denote observational data. A statistical measurement error model is then specified:

$$[Z|Y,\theta_1], \tag{1}$$

where θ_1 represents a collection of parameters. This model can be tailored to a variety of sampling plans. Perhaps the simplest is that we observe, with error, the process of interest (Y) at some sites and times. That is, the data set is $Z=\{Z(s,t):(s,t)\in\mathcal{D}\}$, where $\mathcal{D}\subset\mathcal{R}^2$. A standard example is to assume that, conditional on Y and θ_1, the $Z(s,t)$ are independent and that in each case

$$Z(s,t)\sim Gau(Y(s,t),\sigma^2_{s,t}). \tag{2}$$

In this case, $\theta_1=\{\sigma^2_{s,t}:(s,t)\in\mathcal{D}\}$ is the set of measurement error variances.

More interesting models are possible. In particular, observations $Z\in D\subset\mathcal{R}^2$ need not

coincide with the Y process ($Y \in \mathcal{M}$); we need only construct a probability model (1) for such observations given the gridded values of Y. (The reader should not infer that we believe this step is "simple.") As another important example, note that spatial and/or temporal averaging schemes can often be modeled via linear operations on Y. For example, spatial averaging yields data given by $\mathbf{Z}_i = K\vec{\mathbf{Y}}_t + \mathbf{e}$, where the measurement errors, $\mathbf{e}$, are from some multivariate distribution. In some cases, it may be appropriate to include some unknown parameters in K.

3.2 Second Stage: Large- and Small-Scale Features.

The suggested modeling strategies given below for the process Y are particularly relevant for atmospheric and oceanographic processes. Such processes are expected to display both strong seasonal variations and regional structures.

The model for Y is conditional on three processes, denoted by $\boldsymbol{\mu}, \boldsymbol{\beta}$, and $X = \{X(s,t) : (s,t) \in \mathcal{M}\}$, and a collection of parameters θ_2. We separate large-scale temporal behavior from that at short time scales. Assume that for each site and time point,

$$Y(s,t) = \mu(s) + M(t;\beta(s)) + X(s,t) + \gamma(s,t), \tag{3}$$

where

- $\mu(s)$ represents a site-specific mean. Define $\boldsymbol{\mu} \equiv (\mu(1), \ldots, \mu(S))$.
- $M(t;\beta(s))$ is a large-scale temporal model with site specific parameters, $\beta(s)$. For climate data, we would typically model seasonal or monthly effects by letting M be a "cosine function" or a collection of indicator variables for monthly means. Define $\boldsymbol{\beta} \equiv (\beta(1), \ldots, \beta(S))$.
- the $X(s,t)$'s represent a short-time scale, dynamical process.
- the $\gamma(s,t)$'s are zero-mean random variables that model noise.

The role of the X-process is to account for both spatial and temporal dynamics beyond those accounted for in long term means and seasonal behavior. An appeal to terminology from the atmospheric/oceanographic and allied sciences is illustrative: We roughly view the contributions

to Y modeled through $\boldsymbol{\mu}$ and M as *climatic*, whereas we think of the adjustments to climate, modeled by X, as *weather*.

Once the model for X is specified, attention is turned to the $\gamma(s,t)$'s. These quantities represent the unexplained variations in the second stage. A key question is how much space-time structure is to be modeled through the specification of the joint distribution of the $\gamma(s,t)$'s. In principle, one must model an $ST \times ST$ covariance matrix. However, the point of the hierarchical approach is that modeled features, such as X, explain much of the space-time structure of the Y process, so that one might assume that the $Y(s,t)$'s are all *conditionally* independent random variables. Here, we assume:

$$Y(s,t) \sim N(\mu(s) + M(t;\beta(s)) + X(s,t), \sigma_Y^2(s)), \tag{4}$$

where $\theta_2 = (\sigma_Y^2(1), \ldots, \sigma_Y^2(S))$.

3.3 Third Stage: Spatial Structures and Dynamics.

We assume that $\boldsymbol{\mu}, \boldsymbol{\beta}$, and X are mutually independent, conditional on third-stage parameters, θ_3. We also assume that θ_3 can be partitioned as $\theta_3 = (\theta_\mu, \theta_\beta, \theta_X)$, leading to

$$[\boldsymbol{\mu}, \boldsymbol{\beta}, X|\theta_3] = [\boldsymbol{\mu}|\theta_\mu][\boldsymbol{\beta}|\theta_\beta][X|\theta_X]. \tag{5}$$

To capture gross spatial structures, we envision use of spatial models, such as Markov random fields, for $[\boldsymbol{\mu}|\theta_\mu]$ and $[\boldsymbol{\beta}|\theta_\beta]$.

Physically and/or experimentally based dynamical models for X are possible. A popular and flexible collection of models are (one-step) Markovian time evolution models (e.g., see Berliner, 1996). Formally, we assume that

$$[X|\theta_X] = \prod_t [\mathbf{X}(,t+1)|\mathbf{X}(,t), \theta_X]. \tag{6}$$

Within the class of "statistical" or "stochastic" models, the most common example of (6) is a

(conditional) vector autoregression,

$$\vec{\boldsymbol{X}}_{t+1} = \mathbf{H}\vec{\boldsymbol{X}}_t + \vec{\boldsymbol{\eta}}_{t+1}, \tag{7}$$

where $\mathbf{H}$ is an $S \times S$ matrix of regression coefficients and $\{\vec{\boldsymbol{\eta}}_t\}$ is an independent sequence of zero-mean errors. Issues in the modeling of the $\vec{\boldsymbol{\eta}}_t$'s are similar to those discussed above concerning the $\gamma(s,t)$'s. The crucial issue of modeling spatial structures clearly involves trade-offs between the richness of the matrix $\mathbf{H}$ versus the level of independence among the elements of any given $\vec{\boldsymbol{\eta}}_t$. A standard default model is that the elements of each $\vec{\boldsymbol{\eta}}_t$ are spatially independent, zero-mean Gaussian random variables, each having variance σ_η^2. (In this case $\theta_X = (\mathbf{H}, \sigma_\eta^2)$.)

To clarify the points raised in the previous paragraph, first note that if $\mathbf{H}$ is assumed to be diagonal, then the entire formulation, based on (3) and (7), is a Bayesian seasonal time series model, with autoregressive shocks and spatially dependent parameters. However, we may not have confidence that spatial dependencies are explained sufficiently in such a model and so we might change the model to one with non-diagonal covariance matrix for the shocks $\vec{\boldsymbol{\eta}}_t$. Alternatively, a non-diagonal $\mathbf{H}$ may be a natural choice. For example, we could write models which essentially regress each element of $\vec{\boldsymbol{X}}_{t+1}$ onto elements of $\vec{\boldsymbol{X}}_t$, corresponding to some selection of neighboring sites. (See Cressie, 1993, Section 6.3 for related discussion.) In cases where this strategy is effective, the covariance matrix for the shocks $\vec{\boldsymbol{\eta}}_t$ could be modeled as diagonal, or nearly so.

Higher-order temporal dependencies can be readily modeled. For example, we can certainly write down a k-step vector autoregression,

$$\vec{\boldsymbol{X}}_{t+1} = \sum_{l=1}^{k} \mathbf{H}_l \vec{\boldsymbol{X}}_{t+1-l} + \vec{\boldsymbol{\eta}}_{t+1}. \tag{8}$$

Assumptions are as before with (7) and $\{\mathbf{H}_l : l = 1, \ldots, k\}$ are $S \times S$ matrices of regression coefficients. This is a special case of a class of models known as the space-time autoregressive moving-average (STARMA) models (e.g., see Cressie, 1993, p. 450). However, dimensionality of the specification rapidly becomes a limiting consideration. As a partial solution to the dimensionality problem, additional modeling is possible. For example, in (8), one might introduce various parameterizations of the matrices $\mathbf{H}_k$. At one extreme, all the $\mathbf{H}_k$'s might be assumed to be

diagonal, whose elements could be endowed with spatially dependent priors at the next stage. Again, an interaction between the modeling of the $\mathbf{H}_k$ and the covariance matrix of the noise vectors $\{\vec{\eta}_t\}$ arises.

The Bayesian approach is natural for combining statistical and physical modeling. Operationally, we may imagine two basic strategies. The first is to employ flexible classes of "statistical" models, such as the vector autoregressions discussed earlier, and then add scientific understanding via the specification of the prior on parameters of the model. A second approach is to develop, explicitly, physically based "mechanistic" specifications for use in analogs of (6). We will pursue this latter direction elsewhere.

A final important comment involves "identifiability" issues. Depending on the modeling strategies used in this third stage, the implied models might not be identifiable. Intuitively, the source of the problem is that $X(s,t)$ and $\gamma(s,t)$ appear only through their sum in (3). Without special sampling plans, parameters of the distributions of the X's and γ's cannot be identified. In a Bayesian analysis with *proper* probabilities on all quantities, identifiability issues do not prohibit us from proceeding, though we should be careful in interpretating results for unidentified parameters. See Besag et al. (1995) for general discussion, and our Section 4 for amplification in an example.

3.4 Fourth Stage: Priors on Parameters

The fourth stage is the specification of priors for all model parameters. Our problem is the specification of $[\theta_1, \theta_2, \theta_3|\theta_4]$, θ_4 is some collection of *hyperparameters*. It is convenient to assume a partition, $\theta_4 = (\theta_4(1), \theta_4(2), \theta_4(3))$, into hyperparameters associated with each stage, and a conditional independence relation,

$$[\theta_1, \theta_2, \theta_3|\theta_4] = [\theta_1|\theta_4(1)][\theta_2|\theta_4(2)][\theta_3|\theta_4(3)]. \tag{9}$$

Further, $\theta_4(3)$ would typically be partitioned as $\theta_4(3) = (\theta_4(\mu), \theta_4(\beta), \theta_4(X))$, and coupled with a further conditional independence assumption,

$$[\theta_3|\theta_4(3)] = [\theta_\mu|\theta_4(\mu)][\theta_\beta|\theta_4(\beta)][\theta_X|\theta_4(X)]. \tag{10}$$

3.5 Fifth Stage: Hyperpriors

Finally, hyperparameter priors are specified. The standard assumption is that

$$[\theta_4] = [\theta_4(1)][\theta_4(2)][\theta_4(\mu)][\theta_4(\beta)][\theta_4(X)]. \tag{11}$$

Often, the formulation is simplified by taking $\theta_4(1)$ and/or $\theta_4(2)$ to be either empty or known, so that the corresponding terms on the right hand side of (11) drop out.

4 Example: Monthly Averaged Maximum Temperature

4.1 Data

We demonstrate some of the concepts in our Bayesian hierarchical space-time model with a monthly averaged maximum temperature data set in a portion of the midwest "corn belt". In particular, 20 years (1974-1993) of monthly averaged temperatures at 131 stations from the United States Historical Climatology Network data set (Easterling et al., 1995) are considered. The locations of these stations are shown in Figure 1. We wish to predict monthly temperature (240 months) at $10 \times 7 = 70$ locations on a one degree latitude by one degree longitude grid covering 9 degrees of longitude and 6 degrees of latitude (see Figure 1). Note that there are missing data at many observation sites. Of particular interest is the space-time dynamical structure of the non-seasonal portion of the maximum temperature field.

4.1.1 Exploratory Analysis

In order to get a general understanding of the data, we have performed some simple exploratory analyses. Figure 2 shows a contour plot derived from the temporal means for each observation location. Note that there is a large-scale spatial trend from southwest to northeast. It appears that a simple linear trend surface, including an intercept and a linear longitude and linear latitude term, would be a reasonable large-scale spatial structure model for the data in this case.

Further examination of the data suggest that there is a significant annual cycle in temperature. Figure 3 shows a contour map of the estimates of the "cosine" coefficient of the annual harmonic

for each site after removal of the site-specific mean shown in Figure 2. The annual harmonic for a given site $\mathbf{s}_k$ is given by $m(\mathbf{s}_k, t) = f(\mathbf{s}_k)cos(2\pi t/12) + g(\mathbf{s}_k)sin(2\pi t/12)$, where t is a month index (1 to 240) and $f(\mathbf{s}_k)$ and $g(\mathbf{s}_k)$ are the "cosine" and "sine" coefficients, respectively. Apparently, there is large-scale spatial structure in the "cosine" component, including a linear trend from north to south. Similarly, Figure 4 shows a contour plot of the "sine" coefficients of the annual harmonic. We detect a large-scale spatial linear trend here as well.

After removing the site-specific means and site-specific annual harmonic from the observations, Figure 5 shows a contour plot of the estimated site-specific autoregressive order 1 [AR(1)] parameters. This plot suggests that there is spatial structure in the AR(1) parameters and that some regions have increased temporal predictability (e.g., south-central Minnesota). Figure 6 shows a contour plot of the estimated standard deviation of AR(1) shocks at each site . We note that there is spatial structure to the shocks (i.e., a linear trend from southwest to northeast) and the associated variances exhibit heteroskedasticity.

4.2 Temperature Model

The model for maximum temperature follows the hierarchical framework outlined in Section 3. In the subsections below, we develop each stage in turn.

4.2.1 First Stage: Measurement Process

We are given data $Z(\mathbf{r}_j, t), j = 1, \ldots, m$ at $m = 131$ spatial locations which may or may not be at a grid location $\mathbf{s}_i, i = 1, \ldots, S$ $(S = 70)$. Consider the model,

$$\vec{\boldsymbol{Z}}_t = \mathbf{K}\vec{\boldsymbol{Y}}_t + \vec{\boldsymbol{\epsilon}}_t \tag{12}$$

where $\vec{\boldsymbol{Z}}_t$ is an $m \times 1$ vector of observations, $\vec{\boldsymbol{Y}}_t$ is an $S \times 1$ state vector, $\mathbf{K}$ is an $m \times S$ matrix that maps the temperature values at grid locations to the observations, and $\vec{\boldsymbol{\epsilon}}_t$ is an $m \times 1$ error vector representing sub-grid scale processes and measurement error. In the case of the temperature data, a plot of covariance as a function of distance (Figure 7, bottom panel) suggests that we can assume the sub-grid spatial structure is white and that conditional on the true process $Y(\mathbf{s}, t)$,

the other observations within the .5 degree by .5 degree box surrounding $\mathbf{s}_i$, are i.i.d. Gaussian random variables with mean $Y(\mathbf{s}_i, t)$ and variance σ_ϵ^2 (with $\sigma_\epsilon^2 \approx 2$). Thus, $\mathbf{K}$ is a matrix of zeros and ones, with a one in the $(i,j)^{th}$ position indicating that the i^{th} grid location is associated with the j^{th} observation location. No attempt is made to distinguish between sub-grid scale variance and measurement error in this example. Although plausible, the model described here is rather simple. We recognize that the choice of the appropriate model at this stage is a research problem in itself, but the investigation of such a model is not our primary message in this paper.

4.2.2 Second Stage: Large- and Small-Scale Features

We now consider the model for the state process:

$$\vec{\boldsymbol{Y}}_t = \vec{\boldsymbol{\mu}} + \vec{\boldsymbol{M}}_t + \vec{\boldsymbol{X}}_t + \vec{\boldsymbol{\gamma}}_t \tag{13}$$

where each term is an $S \times 1$ vector, $\vec{\boldsymbol{\mu}}$ are gridpoint means which are assumed to have spatial structure, $\vec{\boldsymbol{M}}_t$ is a gridpoint dependent seasonal term, $\vec{\boldsymbol{X}}_t$ is a dynamical process that accounts for intra-seasonal space-time variability, and $\vec{\boldsymbol{\gamma}}_t$ is an error term. We assume that the $\vec{\boldsymbol{Y}}_t$'s are conditionally Gaussian, such that:

$$[\vec{\boldsymbol{Y}}_t | \vec{\boldsymbol{\mu}}, \vec{\boldsymbol{M}}_t, \vec{\boldsymbol{X}}_t, \sigma_\gamma^2] \sim Gau(\vec{\boldsymbol{\mu}} + \vec{\boldsymbol{M}}_t + \vec{\boldsymbol{X}}_t, \sigma_\gamma^2 \mathbf{I}) \tag{14}$$

4.2.3 Third Stage: Spatial Structures and Dynamics

As discussed in Section 3.3, we assume that $\vec{\boldsymbol{\mu}}, \vec{\boldsymbol{M}}_t$ and $\vec{\boldsymbol{X}}_t$ are mutually independent, conditional on third-stage parameters. We then model $\vec{\boldsymbol{\mu}}$ as a conditionally specified Gaussian MRF with first-order spatial dependence. That is, for grid locations $k = 1, \ldots, 10; l = 1, \ldots, 7$, where k and l are indices corresponding to longitudes $\{98, 97, \ldots, 89\}$ and latitudes $\{39, 40, \ldots, 45\}$, respectively:

$$\begin{aligned} \mu(k,l) | \{\mu(i,j) : (i,j) \neq (k,l)\} \sim\ & Gau(\mu_0(k,l) + \alpha_\mu\{(\mu(k-1,l) - \mu_0(k-1,l)) \\ & + (\mu(k+1,l) - \mu_0(k+1,l))\} + \beta_\mu\{(\mu(k,l-1) - \\ & \mu_0(k,l-1)) + (\mu(k,l+1) - \mu_0(k,l+1))\}, \tau_\mu^2), \end{aligned} \tag{15}$$

where $\vec{\boldsymbol{\mu}}_0$ is the gridpoint specific MRF mean, α_μ and β_μ are zonal (east-west) and meridional (north-south) MRF spatial dependence parameters, respectively, and τ_μ^2 is the homogeneous MRF variance. It can be shown as a result of the Factorization Theorem (e.g., Cressie 1993, p.414) that:

$$\vec{\boldsymbol{\mu}}|\{\vec{\boldsymbol{\mu}}_0, \tau_\mu^2, \alpha_\mu, \beta_\mu\} \sim Gau(\vec{\boldsymbol{\mu}}_0, (\mathbf{I} - \mathbf{C}_\mu)^{-1}\tau_\mu^2), \tag{16}$$

where $\mathbf{C}_\mu$ is an $S \times S$ matrix with four off-diagonals given by α_μ and β_μ (e.g., Cressie 1993, p.434).

The seasonal component is modeled as an annual harmonic with amplitudes and phases that vary spatially:

$$\vec{\boldsymbol{M}}_t = \vec{\boldsymbol{f}}\ cos(\omega t) + \vec{\boldsymbol{g}}\ sin(\omega t) \tag{17}$$

where ω is specified to be $2\pi/12$ for the annual harmonic, and $\vec{\boldsymbol{f}}$ and $\vec{\boldsymbol{g}}$ are spatially varying "cosine" and "sine" coefficients, respectively. We model the "cosine" and "sine" coefficients of the annual harmonic given by (17) as simple spatial trends:

$$f(k,l) = f[1] + f[2]\,\mathrm{long}(k) + f[3]\,\mathrm{lat}(l) \tag{18}$$

$$g(k,l) = g[1] + g[2]\,\mathrm{long}(k) + g[3]\,\mathrm{lat}(l) \tag{19}$$

where $f[1], f[2], f[3]$ and $g[1], g[2], g[3]$ are independent Gaussian random variables:

$$f[1] \sim Gau(\tilde{f}[1], \tilde{\sigma}_f^2[1]) \tag{20}$$

$$f[2] \sim Gau(\tilde{f}[2], \tilde{\sigma}_f^2[2]) \tag{21}$$

$$f[3] \sim Gau(\tilde{f}[3], \tilde{\sigma}_f^2[3]) \tag{22}$$

$$g[1] \sim Gau(\tilde{g}[1], \tilde{\sigma}_g^2[1]) \tag{23}$$

$$g[2] \sim Gau(\tilde{g}[2], \tilde{\sigma}_g^2[2]) \tag{24}$$

$$g[3] \sim Gau(\tilde{g}[3], \tilde{\sigma}_g^2[3]), \tag{25}$$

where the parameters are fixed and specified as in Table 2.

The space-time dynamic term is modeled as a vector autoregressive (VAR) process:

$$\vec{\boldsymbol{X}}_t = \mathbf{H}\vec{\boldsymbol{X}}_{t-1} + \vec{\boldsymbol{\eta}}_t \tag{26}$$

where $\mathbf{H}$ is an $S \times S$ parameter matrix, and $\vec{\boldsymbol{\eta}}_t$ is the VAR noise term. With the large spatial dimensions typical of climatological problems, even this simple VAR model must be further simplified in practice. We consider the "nearest neighbor" VAR model for the parameter matrix. That is,

$$\begin{aligned} X((k,l),t) &= a(k,l)\, X((k,l),t-1) + b\, X((k+1,l),t-1) \\ &\quad + c\, X((k,l+1),t-1) + d\, X((k-1,l),t-1) \\ &\quad + e\, X((k,l-1),t-1) + \eta((k,l),t). \end{aligned} \tag{27}$$

This is a form of the STARMA models described by Cressie (1993, p.450). Note that here the autoregressive parameter for the grid location under consideration, $a(k,l)$, is allowed to vary spatially, but the nearest neighbor parameters do not vary spatially. It is anticipated that estimates of b, c, d, and e will have implications related to the dynamics of the climatological system. To simplify the model further, we assume that $\{\eta((k,l),t)\}$ is spatial white noise. This assumption may be unrealistic (because it assumes no instantaneous spatial interaction), but it is hoped that, through marginalization, the spatial dependence in the $\{\vec{\boldsymbol{a}}\}$ parameters, along with the lagged nearest neighbor structure, will provide suitable spatial dependence in the X process. In addition, large-scale instantaneous spatial interaction is captured in the $\vec{\boldsymbol{\mu}}$ and $\vec{\boldsymbol{M}}_t$ terms. Thus, we assume:

$$\vec{\boldsymbol{X}}_t | \{\vec{\boldsymbol{X}}_{t-1}, \mathbf{H}, \sigma_\eta^2\} \;\sim\; Gau(\mathbf{H}\vec{\boldsymbol{X}}_{t-1}, \sigma_\eta^2 \mathbf{I}), \tag{28}$$

where $\mathbf{H}$ is a diagonal matrix with four off-diagonals given by the parameters b, c, d, e, and a main diagonal containing the vector $\vec{\boldsymbol{a}}$. For edge sites, some of the off-diagonals have corresponding zeros.

A "diagonal" VAR model (e.g., b, c, d, e equal to zero) was also considered, as well as a "spatially dependent" nearest neighbor VAR model (e.g., $\vec{\boldsymbol{b}}, \vec{\boldsymbol{c}}, \vec{\boldsymbol{d}}, \vec{\boldsymbol{e}}$ are allowed to vary spatially). The

former does not capture as much spatial structure in the X field as the model described here. Furthermore, we feel the latter model is overparameterized and it is *very* sensitive to the initial conditions and the associated Gibbs sampler has difficulty converging.

4.2.4 Fourth Stage: Priors on Parameters

As indicated in Section 3.4, we partition the fourth stage priors and assume conditional independence.

We represent the mean $\{\vec{\boldsymbol{\mu}}_0\}$ of the gridpoint mean MRF given by (16) as a simple spatial trend:

$$\mu_0(k,l) = \mu_0[1] + \mu_0[2]\,\mathrm{long}(k) + \mu_0[3]\,\mathrm{lat}(l) \tag{29}$$

where $\mathrm{long}(k)$ and $\mathrm{lat}(l)$ are the longitude and latitude of the $(k,l)^{th}$ gridpoint, respectively, and the regression coefficients $\mu_0[1], \mu_0[2], \mu_0[3]$ are specified to be independent Gaussian random variables:

$$\mu_0[1] \sim Gau(\tilde{\mu}_0[1], \tilde{\sigma}^2_{\mu_0}[1]) \tag{30}$$

$$\mu_0[2] \sim Gau(\tilde{\mu}_0[2], \tilde{\sigma}^2_{\mu_0}[2]) \tag{31}$$

$$\mu_0[3] \sim Gau(\tilde{\mu}_0[3], \tilde{\sigma}^2_{\mu_0}[3]), \tag{32}$$

where the parameters are fixed and specified as in Table 2.

We let the spatial dependence parameters α_μ, β_μ in (15) be independent Gaussian random variables, but constrain them to ensure positive-definiteness of the marginal covariance matrices:

$$\alpha_\mu \sim Gau(\tilde{\alpha_\mu}, \tilde{\sigma}^2_{\alpha_\mu}) \tag{33}$$

$$\beta_\mu \sim Gau(\tilde{\beta_\mu}, \tilde{\sigma}^2_{\beta_\mu}), \tag{34}$$

with fixed and specified parameters as in Table 2.

We represent the spatially varying autoregressive parameters $\vec{\boldsymbol{a}}$, given by (27), as a condition-

ally specified Gaussian MRF with first-order spatial dependence:

$$\begin{aligned} a(k,l)|\{a(i,j):(i,j)\neq(k,l)\} \sim\ & Gau(a_0(k,l)+\alpha_a\{(a(k-1,l)-a_0(k-1,l)) \\ & +(a(k+1,l)-a_0(k+1,l))\}+\beta_a\{(a(k,l-1) \\ & -a_0(k,l-1))+(a(k,l+1)-a_0(k,l+1))\},\tau_a^2) \end{aligned} \tag{35}$$

where $\vec{a}_0$ is the MRF mean, α_a and β_a are the zonal and meridional MRF spatial dependence parameters, and τ_a^2 is the homogeneous variance. The distributions of these parameters are specified in the fifth stage.

We also assume that the nearest neighbor autoregressive parameters are independent Gaussian random variables:

$$b \sim\ Gau(\tilde{b_0},\tilde{\sigma}_{a_0}^2) \tag{36}$$

$$c \sim\ Gau(\tilde{c_0},\tilde{\sigma}_{c_0}^2) \tag{37}$$

$$d \sim\ Gau(\tilde{d_0},\tilde{\sigma}_{d_0}^2) \tag{38}$$

$$e \sim\ Gau(\tilde{e_0},\tilde{\sigma}_{e_0}^2), \tag{39}$$

with fixed and specified parameters as in Table 2.

For the variances specified in stages one through three, we assume independence and use the conjugate priors:

$$\sigma_\epsilon^2 \sim\ IG(\tilde{q}_\epsilon,\tilde{r}_\epsilon) \tag{40}$$

$$\sigma_\gamma^2 \sim\ IG(\tilde{q}_\gamma,\tilde{r}_\gamma) \tag{41}$$

$$\sigma_\eta^2 \sim\ IG(\tilde{q}_\eta,\tilde{r}_\eta) \tag{42}$$

$$\tau_\mu^2 \sim\ IG(\tilde{q}_\mu,\tilde{r}_\mu), \tag{43}$$

where IG refers to the inverse Gamma distribution, and the parameters are fixed and specified as in Table 2.

4.2.5 Fifth Stage: Hyperpriors

As suggested by the exploratory analysis (see Figure 5), we assume that the MRF mean for the spatially varying autoregressive process $\vec{\boldsymbol{a}}$ given by (35) has a large-scale linear spatial trend, decreasing from the northwest to the southeast. Thus, we assume that $\vec{\boldsymbol{a}}_0$ has a simple spatial trend structure:

$$a_0(k,l) = a_0[1] + a_0[2]\,\mathrm{long}(k) + a_0[3]\,\mathrm{lat}(l), \tag{44}$$

where $a_0[1], a_0[2], a_0[3]$ are independent Gaussian random variables:

$$a_0[1] \sim Gau(\tilde{a}_0[1], \tilde{\sigma}^2_{a_0}[1]) \tag{45}$$

$$a_0[2] \sim Gau(\tilde{a}_0[2], \tilde{\sigma}^2_{a_0}[2]) \tag{46}$$

$$a_0[3] \sim Gau(\tilde{a}_0[3], \tilde{\sigma}^2_{a_0}[3]), \tag{47}$$

with fixed and specified parameters as in Table 2. Furthermore, we let the spatial dependence parameters α_a, β_a in (35) be independent Gaussian random variables, but constrain them to ensure positive-definiteness of the marginal covariance matrices:

$$\alpha_a \sim Gau(\tilde{\alpha_a}, \tilde{\sigma}^2_{\alpha_a}) \tag{48}$$

$$\beta_a \sim Gau(\tilde{\beta}_a, \tilde{\sigma}^2_{\beta_a}), \tag{49}$$

with fixed and specified parameters as in Table 2. Finally, we assume that the variance parameter τ_a^2 in (35) is independent of α_a and β_a and use the inverse Gamma conjugate prior:

$$\tau_a^2 \sim IG(\tilde{q}_a, \tilde{r}_a), \tag{50}$$

where the parameters are fixed and suitably specified as in Table 2.

4.3 Bayesian Estimation: Gibbs Sampler

Markov chain Monte Carlo (MCMC) methods have been used in stochastic simulation for over 40 years (e.g., Metropolis et al. 1953, Hastings 1970). In particular, the Gibbs sampler approach

to MCMC has been used extensively in the statistics literature since the pioneering work by Geman and Geman (1984). For a general overview of these methods see Gilks et al. (1996). We have chosen a Gibbs sampler to allow us to determine the relevant posterior distributions for the hierarchical model of temperature given in Section 4.2. Since we have chosen Gaussian distributions with conjugate priors, the derivation and implementation of the full conditional distributions needed for the Gibbs sampler are, with one exception, quite straightforward. (See the Appendix for the derivations.) The exception concerns the full conditional distributions for each of the MRF spatial dependence parameters (i.e., $\alpha_\mu, \beta_\mu, \alpha_a, \beta_a$). As an example, the full conditional distribution for α_μ (see Appendix) is:

$$\begin{aligned}
[\alpha_\mu | \ \cdot \ ] &\propto [\vec{\boldsymbol{\mu}} | \vec{\boldsymbol{\mu}}_0, \alpha_\mu, \beta_\mu, \tau_\mu^2][\alpha_\mu] \\
&\propto \frac{1}{|(\mathbf{I} - \mathbf{C}_\mu)^{-1}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}[(\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)'(\mathbf{I} - \mathbf{C}_\mu)(\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)(\tau_\mu^2)^{-1} + \\
&\quad (\alpha_\mu - \tilde{\alpha}_\mu)^2(\tilde{\sigma}_\mu^2)^{-1}]\} \\
&\stackrel{\text{def}}{=} h(\alpha_\mu),
\end{aligned} \tag{51}$$

where α_μ is on two of the four diagonals of $\mathbf{C}_\mu$. The presence of α_μ in the determinant term makes this a difficult conditional distribution from which to sample. In order to proceed, we employ a Metropolis-Hastings step in the Gibbs sampler [e.g., see Gilks 1996]. First, we note that it is relatively easy to sample from the "pseudo full conditional", which makes use of the pseudolikelihood approach to classical MRF estimation [e.g., Cressie 1993 pp. 461-463]. That is,

$$\begin{aligned}
[\alpha_\mu | \ \cdot \ ]_{\text{pseudo}} &\propto \prod_{i=1}^{n} [\mu(\mathbf{s}_i) | \mu(\mathbf{s}_k), k \neq i; \boldsymbol{\mu}_0, \alpha_\mu, \beta_\mu, \tau_\mu^2][\alpha_\mu] \\
&\stackrel{\text{def}}{=} h^*(\alpha_\mu).
\end{aligned} \tag{52}$$

Then, the Metropolis-Hastings step for generating the k^{th} sample from the full conditional distribution of α_μ is performed as follows:

- Choose a starting value for $\alpha_\mu(0)$ (e.g., see Section 4.3.1),
- Sample $\alpha_\mu(k)$ from $h^*(\alpha_\mu)$,

- Sample U from $U(0,1)$,
- If U $\leq \min[1, \frac{h(\alpha_\mu(k))h^*(\alpha_\mu(k-1))}{h(\alpha_\mu(k-1))h^*(\alpha_\mu(k))}]$, then *accept* $\alpha_\mu(k)$; else, set $\alpha_\mu(k) = \alpha_\mu(k-1)$.

In practice, this step is relatively slow because of the evaluation of the determinant in (51). However, by utilizing sparse matrix technology in the implementation, significant speed improvements can be realized.

4.3.1 Assessing the Gibbs Sampler

While theory implies that the Markov chain is guaranteed to converge to the appropriate stationary distribution, implementation issues arise in practice. One must make choices related to influence of starting values; how long to run the chain before convergence; and how best to monitor the chain and perform the desired estimation. These issues are still the focus of ongoing research (eg., see Gilks et al. 1996). We have taken a rather simple approach in our analysis.

Initially, we ran three pilot simulations (4000 iterations each) with different starting values (one representative of prior means, and the others widely dispersed within each parameter's prior distribution). In addition to a visual assessment of convergence, we examined the Gelman and Rubin (1992) convergence monitor ($\hat{R}$, which should be close to one for convergence) for the model parameters listed in Table 2. Visual assessment of the three pilot simulations suggested that all parameters had "converged" by 1500 iterations. Thus, we discarded the first 1500 iterations and calculated the Gelman and Rubin monitor with the remaining 2500 iterates. As indicated in Table 2, $\sqrt{\hat{R}}$ values for all parameters suggested convergence at 500 iterations (beyond the 1500 burn-in), with monitor values below 1.03 in all cases. We then ran a single long (8000 iterations) chain (based on the prior mean starting value set) and conservatively discarded the 2000 iteration "burn in" period. Posterior means of all quantities were estimated by the corresponding sample means from the long run (neglecting the burn-in period). Due to the correlation of MCMC samples, we estimated the Monte Carlo standard errors by the "batch means" approach described in Roberts (1996). The batch size was determined from examination of lag autocorrelation plots of several parameters as obtained from the pilot samples. In particular, all parameters have autocorrelations that dropped to insignificance by lag 50. Consequently, we used batch sizes of 50

from the remaining 6000 Gibbs iterations. Although we are not guaranteed that the batch means are uncorrelated with this batch size, we believe it to be a conservative size for most parameters in our model. The Monte Carlo parameter estimates from this simulation, as well as the associated Monte Carlo standard errors, are given in Table 2.

From a modeling perspective, the results from the Gibbs sampling and the rate of convergence were very encouraging in general. Note that the large-scale spatial trend parameters $\{f[i],\ g[i],\ \mu[i];\ i = 1,\ldots,3\}$ do not show a large difference between the prior mean and the posterior mean. The prior means and variances were estimated by regression in the exploratory phase. The lack of Bayesian learning suggests that the trend terms were well estimated in the *ad hoc* exploratory analysis. In contrast, note that the posterior means of the variance components σ_γ^2 and σ_η^2 are extremely different from their prior counterparts. While this may indicate Bayesian "learning," we must also recall our earlier discussion regarding identifiability. From a classical perspective, these variance components are not identifiable in our model. However, prior specifications on these variance components permits a Bayesian analysis. In our case, the prior means given to these components were based on exploratory data analysis, and our confidence in them is not as great as is suggested by their relatively narrow prior variances. Furthermore, sensitivity analysis with different priors suggest that the model tends to put more variance in the σ_γ^2 component, but that the remaining parameters are not greatly sensitive to these changes. In any case, we should not try to interpret these parameters separately, though their sum is meaningful. Note that the identifiability issue could have been handled by other methods (e.g., deletion of one of the two components of variance, reparameterization). We choose to keep the model in the form presented here to show the flexibility available in error modeling. For example, in some meteorological examples, it is natural to model "regimes" by constructing a mixture model for one of the two components of variance.

4.3.2 Results

From the posterior parameter estimates, we can examine some climatological issues. For instance, the MRF and VAR parameters provide useful information. The lagged nearest neighbor parameters $\{b, c, d, e\}$ (36)-(39) suggest that the "upwind" (westward) location (i.e., d) is more influential

Table 2: Gibbs Sampler Results

Model Parameter	Prior Mean	Prior Std. Dev.	MCMC Posterior Mean (Std. Error)	Convergence Monitor ($\hat{R}$)
$f[1]$	17	2.4	16.63 (.037)	1.000
$f[2]$	.10	.032	.097 (.0004)	1.001
$f[3]$	-.75	.032	-.76 (.0006)	1.001
$g[1]$	2.0	.71	2.01 (.0099)	1.002
$g[2]$	.025	.01	.024 (.0001)	1.002
$g[3]$	-.35	.032	-.34 (.0004)	.999
"mean($\boldsymbol{\mu}$)"			59.7	
$\mu[1]$	111	3.74	111.0 (.08)	1.001
$\mu[2]$	-.27	.05	-.27 (.0008)	1.005
$\mu[3]$	-1.8	.16	-1.82 (.002)	1.000
α_μ	.15	.2	.11 (.003)	1.000
β_μ	.15	.2	.022 (.005)	1.005
τ_μ^2	1	.32	.66 (.003)	1.001
b	.0	.28	.096 (.0007)	1.007
c	.08	.28	.14 (.0007)	1.001
d	.12	.28	.20 (.0006)	1.004
e	.07	.28	.15 (.0009)	1.000
σ_ϵ^2	2	.32	2.4 (.0005)	1.000
σ_γ^2	.8	.71	11.8 (.02)	1.019
σ_η^2	23	2.6	8.9 (.02)	1.023
"mean($\mathbf{a}$)"			.22	
$a0[1]$	-2.0	.71	-2.10 (.013)	1.002
$a0[2]$	-.015	.084	-.019 (.0001)	1.001
$a0[3]$	.020	.0084	.014 (.0003)	1.002
α_a	.15	.2	.14 (.006)	1.003
β_a	.15	.2	.14 (.006)	1.003
τ_a^2	.005	.01	.0017 (.0002)	1.022
$X[(97,44),10]$			1.4 (.078)	1.005
$X[(96,41),90]$			.3 (.032)	1.008
$X[(94,42),170]$			1.4 (.078)	1.003
$Y[(97,41),50]$			24.8 (.013)	1.000
$Y[(92,45),130]$			54.3 (.11)	1.007
$Y[(91,42),210]$			84.3 (.11)	1.002

than the "downwind" (eastward) location (i.e., b), in agreement with the prevailing westerly (i.e., west to east) propagation of midlatitude weather systems. In contrast, the northward (i.e., c) and southward (i.e., e) locations are equally important, but not as "influential" as location d. Thus, although the zonal (east-west) upwind component is most influential (as expected), there are significant contributions from north-south locations, suggesting that on monthly time scales there are meridionally propagating processes that are influential. An interesting extension to the model presented here could examine if there is seasonal and/or interannual variation in these parameters as well.

The MRF spatial dependence parameters can also be interpreted from a climatological perspective. After removal of the large-scale spatial trend, α_μ (33) is larger than β_μ (34), suggesting the added importance of zonal (east-west) disturbances in the mean process. However, the α_a (48) and β_a (49) parameter estimates and posterior distributions are very similar (see Figure 8), suggesting that the zonal and meridional influences on spatial variability of the $\vec{\boldsymbol{a}}$ process (35) are both important.

It is also instructive to examine the predicted linear trend surfaces in the model. Figure 9 shows the trend surface derived from the relevant posterior means for the mean maximum temperature (29), the annual harmonic cosine (18) and sine (19) terms, and the same-site autoregressive parameters $\vec{\boldsymbol{a}}_0$ (44). Certainly, we could interpret the $\vec{\boldsymbol{a}}$ trend surface from a climatological perspective. The northwest to southeast trend is interesting. It may be that this is indicative of regional predictability differences; that is, the northwest region has more persistence in monthly temperature (i.e., next month's temperature deviation from climatology is likely to be similar to this month's), so monthly forecasts in that area are likely to have more skill. Another possible explanation, although less likely, is that since the exploratory analysis suggests that the AR(1) noise is not homogeneous [Figure 6], it could be that the model has added the spatial trend in the $\vec{\boldsymbol{a}}$ parameters to compensate for the assumption of homogeneous variance. It is also interesting to examine the annual harmonic amplitude ($amp(\mathbf{s}) \equiv [f(\mathbf{s})^2 + g(\mathbf{s})^2]^{0.5}$) and phase ($phs(\mathbf{s}) \equiv \arctan[-g(\mathbf{s})/f(\mathbf{s})]$), as shown in Figure 10. The annual cycle amplitude and phase increase towards the north-northwest, although the phase difference is inconsequential (with the maximum difference accounting for a phase difference of less than two days). As expected, the phase peak

occurs in July (i.e., $cos(2\pi t/12 + phs(\mathbf{s})), t = 0, \ldots, 11$ is a maximum at $t = 6$, corresponding to a maximum in July, given the estimated $phs(\mathbf{s})$). We also expect that the amplitude should increase as we move towards the northwest, as the winters are colder in northwest corner of the grid than the southeast corner, but regional differences between summer maximum temperature are not as large.

Ultimately, we are interested in the prediction of the X (27) and Y (13) processes. First, consider the posterior samples from the Gibbs iteration for the X process at locations that have observations within the gridbox, and locations that did not have the benefit of observations (Figure 11). As expected, the posterior densities of X are "wider" at locations where no data are present. A similar difference between observation and non-observation locations is evident in the samples from the Y process (not shown). It is also instructive to examine the time variation of the predicted X and Y processes at selected sites. Figure 13 shows time-series plots at two sites, one that is directly influenced by data ($[lon, lat] = [94, 43]$), and one that is not ($[94, 42]$). The model's performance on the Y process is difficult to judge since most of the variance is explained by the annual harmonic. However, we note that there is a marked difference between the series influenced by observations and that which is not influenced directly by observations, at least as they relate to cold season variability. The non-observation site time series fails to capture the increased variability that is present in the winter season in the observation influenced series. In fact, this illustrates that we made a poor modeling assumption. It is well-known that maximum temperatures often show more variability in the winter than in the summer (especially in the central regions of continents), thus implying a seasonal structure to the variance. In retrospect, we should have included such a cyclo-stationary variance term in our model (i.e., let $\gamma(\mathbf{s}, t)$ vary seasonally). Such a formulation is easily handled by our hierarchical structure.

Typically, one is interested in the dynamic processes contained in deviations from the "average climate" (i.e., anomalies), which is what we are modeling in the X process. The first thing to note about the time series of the X process in Figure 13, is that there is a remarkable similarity in the time evolution between these sites. Although not shown, this similarity extends throughout most of the prediction domain. We believe this time-dynamic similarity is physical, since the monthly time-scale weather features should exhibit a fair degree of spatial homogeneity. However, it is

certainly the case that the prediction of the X process at the "non-observation" location shows less variance than the time series at "observation" locations. The smaller variance in the predictor at non-observation sites can be explained as follows. With little or no data in the gridbox, the posterior mean (which is the predictor) is "shrunk" more towards the prior mean than if there were a lot of data, in which case the more variable sample mean tends to dominate. It is also possible that some of these differences may be attributable to the assumption that there is no spatial structure in the VAR variance. As shown in Figure 6, there is likely to be coherent spatial structure in this variance.

To further demonstrate the usefulness of this modeling exercise for examining climatological issues, we focus on the space-time evolution of the X process. In particular, we consider maps of the posterior means of the X process in June for the years 1988, 1990, and 1993 (Figure 13) and for the summer season (June, July, August) of 1988 (Figure 14). The years (1988, 1990, and 1993) were chosen because the region of interest was known to have experienced warm/drought, normal, and cool/wet conditions during these summers, respectively. The summer months were chosen to show the space-time dynamic evolution during one of the extreme (hot) years. Figures 13 and 14 show smoothed posterior mean maps for these periods. The model has correctly represented the "known" weather during the extreme years (Figure 13) and captured the space-time dynamic evolution (Figure 14).

5 Discussion

The hierarchical approach described here offers a flexible approach to modeling a large class of environmental space-time processes. These models are not the most general and many extensions and modifications could be explored as discussed previously. In particular, we have only scratched the surface as far as the space-time dynamic process X is concerned. There are many additional parameterizations that could be investigated. For instance, we could allow time-varying parameters in this process along with physically based parameterizations. These approaches are being considered and will be reported elsewhere. As is always the case, the model must be tailored to the problem at hand. For instance, if our goal is increased climatological understanding (as illustrated

above), then the model discussed here is useful. Similarly, if we are interested in predicting at a future time, our approach is useful. However, if we are trying to predict the X process spatially at locations for which we have never had observations, our lagged nearest neighbor implementation may need to be enhanced. Fortunately, within the hierarchical Bayesian framework, the exploration of various complex space-time models is quite straightforward, and shows promise in space-time environmental problems.

Acknowledgments

Support for this research was provided by the NCAR Geophysical Statistics Project, sponsored by the National Science Foundation under Grant DMS93-12686. N. Cressie's research was partially supported by the U.S. Environmental Protection Agency under co-operative agreement CR 822919-01-0 between the U.S. Environmental Protection Agency and Iowa State University.

6 References

Besag, J., Green, P., Higdon, D., and K. Mengersen, 1995: Bayesian computation and stochastic systems (with discussion), *Statistical Science* **10**, 3-66.

Berliner, L.M., 1996: Hierarchical Bayesian time series models. In *Maximum Entropy and Bayesian Methods*, K. Hanson and R. Silver (Eds.), Kluwer Academic Publishers, Dordrecht, 15-22.

Brown, P.J., N.D. Le, and J.V. Zidek, 1994: Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics*, **22**, 489-509.

Cressie, N.A.C., 1993: *Statistics for Spatial Data, Revised Edition*, Wiley, New York, 990pp.

Easterling, D.R., Karl, T.R., Mason, E.H., Hughes, P.Y., and D.P. Bowman, 1995: *United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data (revised 1995).* NDP-0191R3. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Gelfand, A.E., and A.F.M. Smith, 1990: Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Gelfand, A.E., S.K. Ghosh, J.R. Knight, and C.F. Sirmans, 1995: Spatio-temporal modeling of residential sales data. (obtained from Spatial Statistics Preprint Service, http://hp2.niss.rti.org/organization/personnel/granville/).

Gelman, A., 1996: Inference and monitoring convergence. In *Markov Chain Monte Carlo in Practice.* W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (Eds.), Chapman and Hall, London, pp. 131-140.

Gelman, A. and D.B. Rubin, 1992: Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457-472.

Geman, S., and D. Geman, 1984: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.

Gilks, W.R., 1996: Full conditional distributions. In *Markov Chain Monte Carlo in Practice.* W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (Eds.), Chapman and Hall, London, pp.75-88.

Gilks, W.R., Richardson, S., and D.J. Spiegelhalter (Eds.), 1996: *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, 486pp.

Hainsworth, T.J., and K.V. Mardia, 1992: A Markov random field restoration of image sequences. In *Markov Random Fields: theory and applications*, R. Chellappa and A. Jain, eds., Academic Press, London, 409-445.

Handcock, M. S. and J. R. Wallis, 1994: An approach to statistical spatial-temporal modeling of

meteorological fields (with discussion). *Journal of the American Statistical Association.* **86**, 368-390.

Hastings, W.K., 1970: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

Holton, J.R., 1992: *An Introduction to Dynamic Meteorology*, Academic Press, San Diego.

Huang, H.C. and Cressie, N., 1996: Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis* **22**, 159-175.

Hughes, J.P. and P. Guttorp, 1994: Incorporating spatial dependence and atmospheric data in a model of precipitation. *Journal of Applied Meteorology*, **33**, 1503-1515.

Metropolis, N., Rosenbluth, A.W. Rosenbluth, M.N., Teller, A.H., and E. Teller, 1953: Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087-1092.

Roberts, G.O., 1996: Markov chain concepts related to sampling algorithms. In *Markov Chain Monte Carlo in Practice.* W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (Eds.), Chapman and Hall, London, pp. 45-57.

Waller, L.A., B.P. Carlin, H. Xia, and A.E. Gelfand, 1997: Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**, 607-617.

West, M., and J. Harrison, 1989: *Bayesian Forecasting and Dynamic Models.* Springer-Verlag, New York.

Wikle, C.K., 1996: Spatio-temporal statistical models with applications to atmospheric processes. Ph.D. dissertation, Iowa State University.

Appendix: Derivation of Full Conditional Distributions

This section outlines the derivation of the full conditional distributions used in the Gibbs sampler for the example discussed in Section 4. In general, full conditional distributions are determined by writing the joint distribution of all random quantities divided by the appropriate normalizing constant. In hierarchical models, this process is typically simplified due to the various conditional independence assumptions that are often made. In particular, all components of the full joint distribution that do not functionally depend on the quantity "cancel" from the numerator and denominator of the full conditional distribution. The following derivations begin after these simplifications have been considered.

Note that the Gaussian assumptions in our model implementation lead to straightforward derivation of the full conditional distributions, except in one case (see Section 6.8). Although there are several ways to approach such derivations, we repeatedly use a "completing the squares" approach. For example, the full conditional distribution for some parameter vector $\boldsymbol{\theta}$ is

$$[\boldsymbol{\theta} \mid \cdot\,] \;\propto\; \exp\{-\frac{1}{2}[\boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta} - 2\mathbf{B}\boldsymbol{\theta}]\},$$

which, after completing the square, gives

$$\boldsymbol{\theta} \mid \cdot \;\sim\; Gau(\mathbf{A}^{-1}\mathbf{B}', \mathbf{A}^{-1}).$$

The generic notation $[W \mid \cdot\,]$ and $W \mid \cdot$ is used to represent the conditional distribution of W given *all* other random quantities.

6.1 $[\vec{\boldsymbol{Y}}_t|\cdot]$

Recall from Section 4.2.1:

$$[\vec{\boldsymbol{Z}}_t|\mathbf{K}, \vec{\boldsymbol{Y}}_t, \sigma_\epsilon^2] \;\sim\; Gau(\mathbf{K}\vec{\boldsymbol{Y}}_t, \sigma_\epsilon^2\mathbf{I}), \tag{A. 1}$$

and from Eqn. 14:

$$[\vec{\boldsymbol{Y}}_t|\vec{\boldsymbol{\mu}}, \vec{\boldsymbol{M}}_t, \vec{\boldsymbol{X}}_t, \sigma_\gamma^2] \;\sim\; Gau(\vec{\boldsymbol{\mu}} + \vec{\boldsymbol{M}}_t + \vec{\boldsymbol{X}}_t, \sigma_\gamma^2\mathbf{I}). \tag{A. 2}$$

Then,

$$\begin{aligned}
[\vec{\boldsymbol{Y}}_t|\cdot] &\propto [\vec{\boldsymbol{Z}}_t|\mathbf{K},\vec{\boldsymbol{Y}}_t,\sigma_\epsilon^2][\vec{\boldsymbol{Y}}_t|\vec{\boldsymbol{\mu}},\vec{\boldsymbol{M}}_t,\vec{\boldsymbol{X}}_t,\sigma_\gamma^2] \qquad \text{(A. 3)}\\
&\propto \exp\{-\frac{1}{2\sigma_\epsilon^2}(\vec{\boldsymbol{Z}}_t-\mathbf{K}\vec{\boldsymbol{Y}}_t)'(\vec{\boldsymbol{Z}}_t-\mathbf{K}\vec{\boldsymbol{Y}}_t)\}\exp\{-\frac{1}{2\sigma_\gamma^2}(\vec{\boldsymbol{Y}}_t-[\vec{\boldsymbol{\mu}}+\vec{\boldsymbol{M}}_t+\vec{\boldsymbol{X}}_t])'\\
&\quad\times(\vec{\boldsymbol{Y}}_t-[\vec{\boldsymbol{\mu}}+\vec{\boldsymbol{M}}_t+\vec{\boldsymbol{X}}_t])\}\\
&\propto \exp\{-\frac{1}{2}[(\vec{\boldsymbol{Y}}'_t[\frac{1}{\sigma_\epsilon^2}\mathbf{K}'\mathbf{K}+\frac{1}{\sigma_\gamma^2}\mathbf{I}]\vec{\boldsymbol{Y}}_t)-2(\frac{1}{\sigma_\epsilon^2}\vec{\boldsymbol{Z}}'_t\mathbf{K}+\frac{1}{\sigma_\gamma^2}[\vec{\boldsymbol{\mu}}+\vec{\boldsymbol{M}}_t+\vec{\boldsymbol{X}}_t]')\vec{\boldsymbol{Y}}_t)]\},
\end{aligned}$$

for $t=1,\ldots,T$. Thus,

$$\begin{aligned}
\vec{\boldsymbol{Y}}_t|\cdot \quad &\sim \quad Gau[(\frac{1}{\sigma_\epsilon^2}\mathbf{K}'\mathbf{K}+\frac{1}{\sigma_\gamma^2}\mathbf{I})^{-1}(\frac{1}{\sigma_\epsilon^2}\vec{\boldsymbol{Z}}'_t\mathbf{K}+\frac{1}{\sigma_\gamma^2}[\vec{\boldsymbol{\mu}}+\vec{\boldsymbol{M}}_t+\vec{\boldsymbol{X}}_t]')',\\
&\qquad (\frac{1}{\sigma_\epsilon^2}\mathbf{K}'\mathbf{K}+\frac{1}{\sigma_\gamma^2}\mathbf{I})^{-1}], \qquad \text{(A. 4)}
\end{aligned}$$

for $t=1,\ldots,T$.

6.2 $[\vec{\boldsymbol{X}}_t|\cdot]$

Recall the Markovian assumption (Eqn. 6), as well as Eqn. 14. Then,

$$\vec{\boldsymbol{Y}}_t|\vec{\boldsymbol{\mu}},\vec{\boldsymbol{M}}_t,\vec{\boldsymbol{X}}_t,\sigma_\gamma^2 \;\sim\; Gau(\vec{\boldsymbol{\mu}}+\vec{\boldsymbol{M}}_t+\vec{\boldsymbol{X}}_t,\sigma_\gamma^2\mathbf{I}), \qquad \text{(A. 5)}$$

and Eqn. 20 yields,

$$\vec{\boldsymbol{X}}_t|\vec{\boldsymbol{X}}_{t-1},\mathbf{H},\sigma_\eta^2 \;\sim\; Gau(\mathbf{H}\vec{\boldsymbol{X}}_{t-1},\sigma_\eta^2\mathbf{I}). \qquad \text{(A. 6)}$$

Then, for $t=1,\ldots,T-1$,

$$\begin{aligned}
[\vec{\boldsymbol{X}}_t|\cdot] &\propto [\vec{\boldsymbol{Y}}_t|\vec{\boldsymbol{\mu}},\vec{\boldsymbol{M}}_t,\vec{\boldsymbol{X}}_t,\sigma_\gamma^2][\vec{\boldsymbol{X}}_{t+1}|\vec{\boldsymbol{X}}_t,\mathbf{H},\sigma_\eta^2][\vec{\boldsymbol{X}}_t|\vec{\boldsymbol{X}}_{t-1},\mathbf{H},\sigma_\eta^2] \qquad \text{(A. 7)}\\
&\propto \exp\{-\frac{1}{2}[\frac{1}{\sigma_\gamma^2}(\vec{\boldsymbol{Y}}_t-[\vec{\boldsymbol{\mu}}+\vec{\boldsymbol{M}}_t+\vec{\boldsymbol{X}}_t])'(\vec{\boldsymbol{Y}}_t-[\vec{\boldsymbol{\mu}}+\vec{\boldsymbol{M}}_t+\vec{\boldsymbol{X}}_t])+\\
&\quad\frac{1}{\sigma_\eta^2}(\vec{\boldsymbol{X}}_t-\mathbf{H}\vec{\boldsymbol{X}}_{t-1})'(\vec{\boldsymbol{X}}_t-\mathbf{H}\vec{\boldsymbol{X}}_{t-1})+\\
&\quad\frac{1}{\sigma_\eta^2}(\vec{\boldsymbol{X}}_{t+1}-\mathbf{H}\vec{\boldsymbol{X}}_t)'(\vec{\boldsymbol{X}}_{t+1}-\mathbf{H}\vec{\boldsymbol{X}}_t)]\}\\
&\propto \exp\{-\frac{1}{2}[\vec{\boldsymbol{X}}'_t(\frac{1}{\sigma_\gamma^2}\mathbf{I}+\frac{1}{\sigma_\eta^2}\mathbf{H}'\mathbf{H}+\frac{1}{\sigma_\eta^2}\mathbf{I})\vec{\boldsymbol{X}}_t-2(\frac{1}{\sigma_\gamma^2}(\vec{\boldsymbol{Y}}_t-\vec{\boldsymbol{\mu}}-\vec{\boldsymbol{M}}_t)'\\
&\quad+\frac{1}{\sigma_\eta^2}\vec{\boldsymbol{X}}'_{t+1}\mathbf{H}+\frac{1}{\sigma_\eta^2}\vec{\boldsymbol{X}}'_{t-1}\mathbf{H}')\vec{\boldsymbol{X}}_t]\},
\end{aligned}$$

where we need the initial condition $\vec{\boldsymbol{X}}_0$. Thus,

$$\begin{aligned}\vec{\boldsymbol{X}}_t|\cdot \quad &\sim \quad Gau[(\frac{1}{\sigma_\gamma^2}\mathbf{I}+\frac{1}{\sigma_\eta^2}\mathbf{H}'\mathbf{H}+\frac{1}{\sigma_\eta^2}\mathbf{I})^{-1}(\frac{1}{\sigma_\gamma^2}(\vec{\boldsymbol{Y}}_t-\vec{\boldsymbol{\mu}}-\vec{\boldsymbol{M}}_t)'+\frac{1}{\sigma_\eta^2}\vec{\boldsymbol{X}}'_{t+1}\mathbf{H}+\frac{1}{\sigma_\eta^2}\vec{\boldsymbol{X}}'_{t-1}\mathbf{H}')', \\ &\quad (\frac{1}{\sigma_\gamma^2}\mathbf{I}+\frac{1}{\sigma_\eta^2}\mathbf{H}'\mathbf{H}+\frac{1}{\sigma_\eta^2}\mathbf{I})^{-1}]. \qquad \text{(A. 8)}\end{aligned}$$

Similarly, for $t=T$,

$$[\vec{\boldsymbol{X}}_T|\cdot\,] \quad \propto \quad [\vec{\boldsymbol{Y}}_T|\vec{\boldsymbol{\mu}},\vec{\boldsymbol{M}}_T,\vec{\boldsymbol{X}}_T,\sigma_\gamma^2][\vec{\boldsymbol{X}}_T|\vec{\boldsymbol{X}}_{T-1},\mathbf{H},\sigma_\eta^2], \qquad \text{(A. 9)}$$

which, as above, leads to

$$\begin{aligned}\vec{\boldsymbol{X}}_T|\cdot \quad &\sim \quad Gau[(\frac{1}{\sigma_\gamma^2}\mathbf{I}+\frac{1}{\sigma_\eta^2}\mathbf{I})^{-1}(\frac{1}{\sigma_\gamma^2}(\vec{\boldsymbol{Y}}_T-\vec{\boldsymbol{\mu}}-\vec{\boldsymbol{M}}_T)'+\frac{1}{\sigma_\eta^2}\vec{\boldsymbol{X}}'_{T-1}\mathbf{H}')', \\ &\quad (\frac{1}{\sigma_\gamma^2}\mathbf{I}+\frac{1}{\sigma_\eta^2}\mathbf{I})^{-1}]. \qquad \text{(A. 10)}\end{aligned}$$

Similarly, if $\vec{\boldsymbol{X}}_0|\vec{\boldsymbol{\mu}}_{X_0},\boldsymbol{\Sigma}_{X_0} \sim Gau(\vec{\boldsymbol{\mu}}_{X_0},\boldsymbol{\Sigma}_{X_0})$, the full conditional distribution for $\vec{\boldsymbol{X}}_0$ is given by

$$\begin{aligned}[\vec{\boldsymbol{X}}_0|\cdot\,] \quad &\propto \quad [\vec{\boldsymbol{X}}_1|\vec{\boldsymbol{X}}_0,\mathbf{H},\sigma_\eta^2][\vec{\boldsymbol{X}}_0|\vec{\boldsymbol{\mu}}_{X_0},\boldsymbol{\Sigma}_{X_0}] \qquad \text{(A. 11)} \\ &\propto \quad \exp\{-\frac{1}{2\sigma_\eta^2}(\vec{\boldsymbol{X}}'_0\mathbf{H}'\mathbf{H}\vec{\boldsymbol{X}}_0-2\vec{\boldsymbol{X}}'_1\mathbf{H}\vec{\boldsymbol{X}}_0)-\frac{1}{2}(\vec{\boldsymbol{X}}_0-\vec{\boldsymbol{\mu}}_{X_0})'\boldsymbol{\Sigma}_{X_0}^{-1}(\vec{\boldsymbol{X}}_0-\vec{\boldsymbol{\mu}}_{X_0})\},\end{aligned}$$

which, as above, leads to

$$\begin{aligned}\vec{\boldsymbol{X}}_0|\cdot \quad &\sim \quad Gau[(\frac{1}{\sigma_\eta^2}\mathbf{H}'\mathbf{H}+\boldsymbol{\Sigma}_{X_0}^{-1})^{-1}(\frac{1}{\sigma_\eta^2}\vec{\boldsymbol{X}}'_1\mathbf{H}+\vec{\boldsymbol{\mu}}'_{X_0}\boldsymbol{\Sigma}_{X_0}^{-1})', \\ &\quad (\frac{1}{\sigma_\eta^2}\mathbf{H}'\mathbf{H}+\boldsymbol{\Sigma}_{X_0}^{-1})^{-1}]. \qquad \text{(A. 12)}\end{aligned}$$

6.3 $[\vec{\boldsymbol{\mu}}|\cdot\,]$

Using the distributions in Eqns. 14 and 16, we can derive

$$\begin{aligned}[\vec{\boldsymbol{\mu}}|\cdot\,] \quad &\propto \quad [\vec{\boldsymbol{\mu}}|\vec{\boldsymbol{\mu}}_0,\tau_\mu^2,\alpha_\mu,\beta_\mu]\prod_{t=1}^{T}[\vec{\boldsymbol{Y}}_t|\vec{\boldsymbol{\mu}},\vec{\boldsymbol{M}}_t,\vec{\boldsymbol{X}}_t,\sigma_\gamma^2] \qquad \text{(A. 13)} \\ &\propto \quad \exp\{-\frac{1}{2\tau_\mu^2}(\vec{\boldsymbol{\mu}}-\vec{\boldsymbol{\mu}}_0)'(\mathbf{I}-\mathbf{C}_\mu)(\vec{\boldsymbol{\mu}}-\vec{\boldsymbol{\mu}}_0)\} \\ &\quad \times\ \exp\{-\frac{1}{2\sigma_\gamma^2}\sum_{t=1}^{T}(\vec{\boldsymbol{Y}}_t-\vec{\boldsymbol{\mu}}-\vec{\boldsymbol{M}}_t-\vec{\boldsymbol{X}}_t)'(\vec{\boldsymbol{Y}}_t-\vec{\boldsymbol{\mu}}-\vec{\boldsymbol{M}}_t-\vec{\boldsymbol{X}}_t)\}\end{aligned}$$

$$\propto \quad \exp\{-\frac{1}{2}[\vec{\mu}'(\frac{1}{\tau_\mu^2}(\mathbf{I}-\mathbf{C}_\mu)+\frac{T}{\sigma_\gamma^2}\mathbf{I})\vec{\mu}-2(\frac{1}{\tau_\mu^2}\vec{\mu}_0'(\mathbf{I}-\mathbf{C}_\mu)+\frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}(\vec{\boldsymbol{Y}}_t-\vec{\boldsymbol{M}}_t-\vec{\boldsymbol{X}}_t)')\vec{\mu}]\}.$$

Thus,

$$\begin{aligned}\vec{\mu}|\cdot \quad &\sim \quad Gau[(\frac{1}{\tau_\mu^2}(\mathbf{I}-\mathbf{C}_\mu)+\frac{T}{\sigma_\gamma^2}\mathbf{I})^{-1}(\frac{1}{\tau_\mu^2}\vec{\mu}_0'(\mathbf{I}-\mathbf{C}_\mu)+\frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}(\vec{\boldsymbol{Y}}_t-\vec{\boldsymbol{M}}_t-\vec{\boldsymbol{X}}_t)')', \\ &\quad (\frac{1}{\tau_\mu^2}(\mathbf{I}-\mathbf{C}_\mu)+\frac{T}{\sigma_\gamma^2}\mathbf{I})^{-1}]. \qquad \text{(A. 14)}\end{aligned}$$

6.4 $[\vec{\mu}_0|\cdot]$ and $[\vec{a}_0\,|\,\cdot]$

Recall from Eqn. 21 that,

$$\mu_0(k,l)=\mu_0[1]+\mu_0[2]\,\mathrm{long}(k)+\mu_0[3]\,\mathrm{lat}(l), \qquad \text{(A. 15)}$$

where $\mathrm{long}(k)$ and $\mathrm{lat}(l)$ are the longitude and latitude of the $(k,l)^{th}$ gridpoint, respectively, and the regression coefficients $\mu_0[1], \mu_0[2], \mu_0[3]$ are specified to be independent Gaussian random variables as shown in Eqns. 22-24. For all k and l defined for our model, we can write

$$\begin{aligned}\vec{\mu}_0 &= (\mathbf{1}\ \vec{\mathrm{long}}\ \vec{\mathrm{lat}})(\mu_0[1],\ \mu_0[2],\ \mu_0[3])' \\ &\equiv \mathbf{P}\boldsymbol{\mu}_{0L}, \qquad \text{(A. 16)}\end{aligned}$$

where $\mathbf{P}$ is the $S\times 3$ linear trend "design" matrix, and $\boldsymbol{\mu}_{0L}\equiv(\mu_0[1],\ \mu_0[2],\ \mu_0[3])'$ is the linear trend parameter vector. Then, from independence and Eqns. 22 - 24,

$$\boldsymbol{\mu}_{0L}|\tilde{\boldsymbol{\mu}}_{0L},\tilde{\boldsymbol{\Sigma}}_{\mu_0} \ \sim\ Gau(\tilde{\boldsymbol{\mu}}_{0L},\tilde{\boldsymbol{\Sigma}}_{\mu_0}), \qquad \text{(A. 17)}$$

where $\tilde{\boldsymbol{\Sigma}}_{\mu_0}$ is a 3×3 diagonal matrix with $\tilde{\sigma}^2_{\mu_0}[1]$, $\tilde{\sigma}^2_{\mu_0}[2]$, and $\tilde{\sigma}^2_{\mu_0}[3]$ on the main diagonal. Then, recalling Eqn. 16,

$$\begin{aligned}[\boldsymbol{\mu}_{0L}|\cdot] &\propto [\vec{\mu}|\vec{\boldsymbol{\mu}}_0,\alpha_\mu,\beta_\mu,\tau_\mu^2][\boldsymbol{\mu}_{0L}|\tilde{\boldsymbol{\mu}}_{0L},\tilde{\boldsymbol{\Sigma}}_{\mu_0}] \qquad \text{(A. 18)} \\ &\propto \exp\{-\frac{1}{2\tau_\mu^2}(\vec{\mu}-\mathbf{P}\boldsymbol{\mu}_{0L})'(\mathbf{I}-\mathbf{C}_\mu)(\vec{\mu}-\mathbf{P}\boldsymbol{\mu}_{0L})\}\exp\{-\frac{1}{2}(\boldsymbol{\mu}_{0L}-\tilde{\boldsymbol{\mu}}_{0L})'\tilde{\boldsymbol{\Sigma}}_{\mu_0}^{-1}(\boldsymbol{\mu}_{0L}-\tilde{\boldsymbol{\mu}}_{0L})\} \\ &\propto \exp\{-\frac{1}{2}[\boldsymbol{\mu}_{0L}'(\frac{1}{\tau_\mu^2}\mathbf{P}'(\mathbf{I}-\mathbf{C}_\mu)\mathbf{P}+\tilde{\boldsymbol{\Sigma}}_{\mu_0}^{-1})\boldsymbol{\mu}_{0L}-2(\frac{1}{\tau_\mu^2}\vec{\mu}'(\mathbf{I}-\mathbf{C}_\mu)\mathbf{P}+\tilde{\boldsymbol{\mu}}_{0L}'\tilde{\boldsymbol{\Sigma}}_{\mu_0}^{-1})\boldsymbol{\mu}_{0L}]\},\end{aligned}$$

and,

$$\boldsymbol{\mu}_{0L}|\cdot \;\sim\; Gau[(\frac{1}{\tau_\mu^2}\mathbf{P}'(\mathbf{I}-\mathbf{C}_\mu)\mathbf{P}+\tilde{\boldsymbol{\Sigma}}_{\mu_0}^{-1})^{-1}(\frac{1}{\tau_\mu^2}\vec{\boldsymbol{\mu}}'(\mathbf{I}-\mathbf{C}_\mu)\mathbf{P}+\tilde{\boldsymbol{\mu}}_{0L}'\tilde{\boldsymbol{\Sigma}}_{\mu_0}^{-1})', (\frac{1}{\tau_\mu^2}\mathbf{P}'(\mathbf{I}-\mathbf{C}_\mu)\mathbf{P}+\tilde{\boldsymbol{\Sigma}}_{\mu_0}^{-1})^{-1}]. \quad \text{(A. 19)}$$

Note that the derivation of the full conditional distribution for $\vec{\boldsymbol{\alpha}}_0$ is analogous to that for $\vec{\boldsymbol{\mu}}_0$, as presented here.

6.5 $[\vec{\boldsymbol{f}}|\cdot]$ and $[\vec{\boldsymbol{g}}|\cdot]$

From Eqn. 25 and from the definition of the linear "design" matrix $\mathbf{P}$ (Eqn. A.16) we can write:

$$\vec{\boldsymbol{f}} = \mathbf{P}\mathbf{f}_L, \quad \text{(A. 20)}$$

where $\mathbf{f}_L \equiv (f[1],\ \ f[2],\ \ f[3])'$. Also, from Eqns. 27-29,

$$\mathbf{f}_L|\tilde{\mathbf{f}}_L, \tilde{\boldsymbol{\Sigma}}_f \;\sim\; Gau(\tilde{\mathbf{f}}_L, \tilde{\boldsymbol{\Sigma}}_f), \quad \text{(A. 21)}$$

where $\tilde{\mathbf{f}}_L \equiv (\tilde{f}[1],\ \ \tilde{f}[2],\ \ \tilde{f}[3])'$ and $\tilde{\boldsymbol{\Sigma}}_f$ is a 3 by 3 diagonal matrix with $\tilde{\sigma}_f^2[1]$, $\tilde{\sigma}_f^2[2]$, and $\tilde{\sigma}_f^2[3]$ on the main diagonal. Then,

$$\begin{aligned}
[\mathbf{f}_L|\cdot] \;&\propto\; [\mathbf{f}_L|\tilde{\mathbf{f}}_L, \tilde{\boldsymbol{\Sigma}}_f]\prod_{t=1}^{T}[\vec{\boldsymbol{Y}}_t|\vec{\boldsymbol{\mu}}, \vec{\boldsymbol{X}}_t, \vec{\boldsymbol{f}}, \vec{\boldsymbol{g}}, \sigma_\gamma^2] \quad \text{(A. 22)}\\
&\propto\; \exp\{-\frac{1}{2}(\mathbf{f}_L-\tilde{\mathbf{f}}_L)'\tilde{\boldsymbol{\Sigma}}_f^{-1}(\mathbf{f}_L-\tilde{\mathbf{f}}_L)\}\exp\{-\frac{1}{2\sigma_\gamma^2}\sum_{t=1}^{T}[\vec{\boldsymbol{Y}}_t-(\vec{\boldsymbol{\mu}}+\mathbf{P}\mathbf{f}_L\cos\omega t+\mathbf{P}\mathbf{g}_L\sin\omega t+\vec{\boldsymbol{X}}_t)]'\\
&\quad\times[\vec{\boldsymbol{Y}}_t-(\vec{\boldsymbol{\mu}}+\mathbf{P}\mathbf{f}_L\cos\omega t+\mathbf{P}\mathbf{g}_L\sin\omega t+\vec{\boldsymbol{X}}_t)]\}\\
&\propto\; \exp\{-\frac{1}{2}[\mathbf{f}_L'(\tilde{\boldsymbol{\Sigma}}_f^{-1}+\frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}(\cos\omega t)^2\mathbf{P}'\mathbf{P})\mathbf{f}_L-2(\tilde{\mathbf{f}}_L'\tilde{\boldsymbol{\Sigma}}_f^{-1}+\\
&\quad\frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}[\vec{\boldsymbol{Y}}_t-(\vec{\boldsymbol{\mu}}+\mathbf{P}\mathbf{g}_L\sin\omega t+\vec{\boldsymbol{X}}_t)]'\mathbf{P}\cos\omega t)\mathbf{f}_L]\},
\end{aligned}$$

and,

$$\begin{aligned}
\mathbf{f}_L\;|\cdot \;\sim\; &Gau[(\tilde{\boldsymbol{\Sigma}}_f^{-1}+\frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}(\cos\omega t)^2\mathbf{P}'\mathbf{P})^{-1}(\tilde{\mathbf{f}}_L'\tilde{\boldsymbol{\Sigma}}_f^{-1}+\frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}[\vec{\boldsymbol{Y}}_t-(\vec{\boldsymbol{\mu}}+\mathbf{P}\mathbf{g}_L\sin\omega t+\vec{\boldsymbol{X}}_t)]'\mathbf{P}\cos\omega t)',\\
&(\tilde{\boldsymbol{\Sigma}}_f^{-1}+\frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}(\cos\omega t)^2\mathbf{P}'\mathbf{P})^{-1}]. \quad \text{(A. 23)}
\end{aligned}$$

Similarly,

$$\mathbf{g}_L \mid \cdot \quad \sim \quad Gau[(\tilde{\mathbf{\Sigma}}_g^{-1} + \frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}(\sin\omega t)^2\mathbf{P}'\mathbf{P})^{-1}(\tilde{\mathbf{g}}_L'\tilde{\mathbf{\Sigma}}_g^{-1} + \frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}[\vec{\boldsymbol{Y}}_t - (\vec{\boldsymbol{\mu}} + \mathbf{P}\mathbf{f}_L\cos\omega t + \vec{\boldsymbol{X}}_t)]'\mathbf{P}\sin\omega t)',$$
$$(\tilde{\mathbf{\Sigma}}_g^{-1} + \frac{1}{\sigma_\gamma^2}\sum_{t=1}^{T}(\sin\omega t)^2\mathbf{P}'\mathbf{P})^{-1}], \tag{A. 24}$$

where $\mathbf{g}_L \equiv (g[1],\ \ g[2],\ \ g[3])'$ and

$$\mathbf{g}_L|\tilde{\mathbf{g}}_L, \tilde{\mathbf{\Sigma}}_g \ \sim \ Gau(\tilde{\mathbf{g}}_L, \tilde{\mathbf{\Sigma}}_g). \tag{A. 25}$$

6.6 $[\vec{\boldsymbol{a}}|\ \cdot\ ]$

From the lagged-nearest neighbor VAR structure (Eqn. 19), we can write the following decomposition:

$$\begin{aligned}
\mathbf{H}\vec{\boldsymbol{X}}_{t-1} &= \mathrm{diag}(\vec{\boldsymbol{a}})\vec{\boldsymbol{X}}_{t-1} + \mathbf{H}_a\vec{\boldsymbol{X}}_{t-1} \qquad \text{(A. 26)}\\
&= \mathrm{diag}(\vec{\boldsymbol{X}}_{t-1})\vec{\boldsymbol{a}} + \mathbf{H}_a\vec{\boldsymbol{X}}_{t-1}\\
&= \mathbf{X}_{t-1}\vec{\boldsymbol{a}} + \mathbf{H}_a\vec{\boldsymbol{X}}_{t-1},
\end{aligned}$$

where $\mathbf{H}_a$ is the $\mathbf{H}$ matrix with the main diagonal (which is $\vec{\boldsymbol{a}}$) replaced by zeros, and $\mathbf{X}_{t-1} \equiv \mathrm{diag}(\vec{\boldsymbol{X}}_{t-1})$. Then, using Eqns. 20 and 33,

$$\begin{aligned}
[\vec{\boldsymbol{a}} \mid \cdot\ ] &\propto \prod_{t=1}^{T}[\vec{\boldsymbol{X}}_t|\vec{\boldsymbol{X}}_{t-1}, \mathbf{H}, \sigma_\eta^2][\vec{\boldsymbol{a}}|\vec{\boldsymbol{a}}_0, \alpha_a, \beta_a, \tau_a^2] \qquad \text{(A. 27)}\\
&\propto \exp\{-\frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}(\vec{\boldsymbol{X}}_t - \mathbf{X}_{t-1}\vec{\boldsymbol{a}} - \mathbf{H}_a\vec{\boldsymbol{X}}_{t-1})'(\vec{\boldsymbol{X}}_t - \mathbf{X}_{t-1}\vec{\boldsymbol{a}} - \mathbf{H}_a\vec{\boldsymbol{X}}_{t-1})\}\\
&\quad \times \exp\{-\frac{1}{2\tau_a^2}(\vec{\boldsymbol{a}} - \vec{\boldsymbol{a}}_0)'(\mathbf{I} - \mathbf{C}_a)(\vec{\boldsymbol{a}} - \vec{\boldsymbol{a}}_0)\}\\
&\propto \exp\{-\frac{1}{2}[\vec{\boldsymbol{a}}'(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\mathbf{X}_{t-1}\mathbf{X}_{t-1} + \frac{1}{\tau_a^2}(\mathbf{I} - \mathbf{C}_a))\vec{\boldsymbol{a}} -\\
&\quad 2[\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}(\vec{\boldsymbol{X}}_t - \mathbf{H}_a\vec{\boldsymbol{X}}_{t-1})'\mathbf{X}_{t-1} + \frac{1}{\tau_a^2}\vec{\boldsymbol{a}}_0(\mathbf{I} - \mathbf{C}_a)]\vec{\boldsymbol{a}}]\},
\end{aligned}$$

and

$$\vec{\boldsymbol{a}} \mid \cdot \quad \sim \quad Gau[(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\mathbf{X}_{t-1}\mathbf{X}_{t-1} + \frac{1}{\tau_a^2}(\mathbf{I} - \mathbf{C}_a))^{-1}[\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}(\vec{\boldsymbol{X}}_t - \mathbf{H}_a\vec{\boldsymbol{X}}_{t-1})'\mathbf{X}_{t-1} + \frac{1}{\tau_a^2}\vec{\boldsymbol{a}}_0(\mathbf{I} - \mathbf{C}_a)]',$$

$$(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\mathbf{X}_{t-1}\mathbf{X}_{t-1}+\frac{1}{\tau_a^2}(\mathbf{I}-\mathbf{C}_a))^{-1}]. \tag{A. 28}$$

6.7 $[b \mid \cdot\,]$,$[c \mid \cdot\,]$,$[d \mid \cdot\,]$,$[e \mid \cdot\,]$

From the lagged nearest neighbor structure (Eqn. 19) we can write the decomposition:

$$\mathbf{H}\vec{\boldsymbol{X}}_{t-1} = b\vec{\boldsymbol{X}}^{\,b}_{t-1} + \mathbf{H}_b\vec{\boldsymbol{X}}_{t-1}, \tag{A. 29}$$

where $\mathbf{H}_b$ is the $\mathbf{H}$ matrix with the diagonal corresponding to the b coefficient replaced by zeros, and $\vec{\boldsymbol{X}}^{\,b}_{t-1} \equiv \mathbf{J}_b\vec{\boldsymbol{X}}_{t-1}$, where $\mathbf{J}_b$ is an $S \times S$ matrix with ones along the diagonal corresponding to the b coefficient in $\mathbf{H}$, and with zeros elsewhere. The exact form of this matrix is a function of the chosen vectorization and the method for handling edge effects. Then, using Eqns. 20 and 34,

$$\begin{aligned}
[b \mid \cdot\,] &\propto \prod_{t=1}^{T}[\vec{\boldsymbol{X}}_t|\vec{\boldsymbol{X}}_{t-1},\mathbf{H},\sigma_\eta^2][b|\tilde{b}_0,\tilde{\sigma}_b^2] \qquad (A.\ 30)\\
&\propto \exp\{-\frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}(\vec{\boldsymbol{X}}_t - b\vec{\boldsymbol{X}}^{\,b}_{t-1} - \mathbf{H}_b\vec{\boldsymbol{X}}_{t-1})'(\vec{\boldsymbol{X}}_t - b\vec{\boldsymbol{X}}^{\,b}_{t-1} - \mathbf{H}_b\vec{\boldsymbol{X}}_{t-1})\}\\
&\quad \times \exp\{-\frac{1}{2\tilde{\sigma}_b^2}(b-\tilde{b}_0)^2\}\\
&\propto \exp\{-\frac{1}{2}[b(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\vec{\boldsymbol{X}}^{\,b\prime}_{t-1}\vec{\boldsymbol{X}}_{t-1}+\frac{1}{\tilde{\sigma}_b^2})b - 2b(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}(\vec{\boldsymbol{X}}_t-\mathbf{H}_b\vec{\boldsymbol{X}}_{t-1})'\vec{\boldsymbol{X}}^{\,b}_{t-1}+\frac{\tilde{b}_0}{\tilde{\sigma}_b^2})]\},
\end{aligned}$$

and

$$\begin{aligned}
b \mid \cdot \;\sim\; & Gau[(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\vec{\boldsymbol{X}}^{\,b\prime}_{t-1}\vec{\boldsymbol{X}}_{t-1}+\frac{1}{\tilde{\sigma}_b^2})^{-1}(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}(\vec{\boldsymbol{X}}_t-\mathbf{H}_b\vec{\boldsymbol{X}}_{t-1})'\vec{\boldsymbol{X}}^{\,b}_{t-1}+\frac{\tilde{b}_0}{\tilde{\sigma}_b^2}),\\
& (\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\vec{\boldsymbol{X}}^{\,b\prime}_{t-1}\vec{\boldsymbol{X}}_{t-1}+\frac{1}{\tilde{\sigma}_b^2})^{-1}]. \qquad (A.\ 31)
\end{aligned}$$

Note that analogous distributions are obtained for c, d, and e.

6.8 $[\alpha_\mu \mid \cdot\,]$,$[\beta_\mu \mid \cdot\,]$,$[\alpha_a \mid \cdot\,]$,$[\beta_a \mid \cdot\,]$

From Eqns. 16 and 38, and with independence assumptions on the hyperparameters,

$$[\alpha_\mu|\ \cdot\ ] \;\propto\; [\vec{\boldsymbol{\mu}}|\vec{\boldsymbol{\mu}}_0,\alpha_\mu,\beta_\mu,\tau_\mu^2][\alpha_\mu|\tilde{\alpha}_\mu,\tilde{\sigma}^2_{\alpha_\mu}] \tag{A. 32}$$

$$\propto \frac{1}{|(\mathbf{I}-\mathbf{C}_\mu)^{-1}|^{\frac{1}{2}}}\exp\{-\frac{1}{2\tau_\mu^2}[(\vec{\boldsymbol{\mu}}-\vec{\boldsymbol{\mu}}_0)'(\mathbf{I}-\mathbf{C}_\mu)(\vec{\boldsymbol{\mu}}-\vec{\boldsymbol{\mu}}_0)\}$$
$$\times \exp\{-\frac{1}{2\tilde{\sigma}^2_{\alpha_\mu}}(\alpha_\mu-\tilde{\alpha}_\mu)^2\}$$

where α_μ is on two of the four diagonals of $\mathbf{C}_\mu$. The presence of α_μ in the determinant term makes this a difficult conditional distribution from which to sample. In order to proceed, consider the "pseudo conditional distribution", which makes use of the pseudolikelihood approach to classical MRF estimation [e.g., Cressie 1993 pp. 461-463]. That is,

$$[\alpha_\mu|\ \cdot\ ]_{\text{pseudo}} \propto \prod_{i=1}^{n}[\mu(\mathbf{s}_i)|\mu(\mathbf{s}_k), k\neq i; \boldsymbol{\mu}_0, \alpha_\mu, \beta_\mu, \tau_\mu^2][\alpha_\mu|\tilde{\alpha}_\mu, \tilde{\sigma}^2_{\alpha_\mu}] \tag{A. 33}$$
$$\propto \exp\{-\frac{1}{2\tau_\mu^2}[\sum_k\sum_l[\mu(k,l)-\mu_0(k,l)-\alpha_\mu\{\mu(k-1,l)-\mu_0(k-1,l)$$
$$+\mu(k+1,l)-\mu_0(k+1,l)\}-\beta_\mu\{\mu(k,l-1)$$
$$-\mu_0(k,l-1)+\mu(k,l+1)-\mu_0(k,l+1)\}]^2]\}$$
$$\times\ \exp\{-\frac{1}{2\tilde{\sigma}^2_{\alpha_\mu}}(\alpha_\mu-\tilde{\alpha}_\mu)^2\}$$
$$\propto \exp\{-\frac{1}{2}[\alpha_\mu^2(\frac{1}{\tau_\mu^2}\sum_k\sum_l\{\mu(k-1,l)-\mu_0(k-1,l)+\mu(k+1,l)-\mu_0(k+1,l)\}^2+\frac{1}{\tilde{\sigma}^2_{\alpha_\mu}})$$
$$-2\alpha_\mu(\frac{\tilde{\alpha}_\mu}{\tilde{\sigma}^2_{\alpha_\mu}}+\sum_k\sum_l\{\mu(k-1,l)-\mu_0(k-1,l)+\mu(k+1,l)-\mu_0(k+1,l)\}$$
$$\times[\mu(k,l)-\mu_0(k,l)-\beta_\mu\{\mu(k,l-1)-\mu_0(k,l-1)+\mu(k,l+1)-\mu_0(k,l+1)\}])]\}.$$

Then,

$$\alpha_{\mu_{pseudo}}|\cdot \sim Gau[(\frac{1}{\tau_\mu^2}\sum_k\sum_l\{\mu(k-1,l)-\mu_0(k-1,l)+\mu(k+1,l)-\mu_0(k+1,l)\}^2+\frac{1}{\tilde{\sigma}^2_{\alpha_\mu}})^{-1}$$
$$\times(\frac{\tilde{\alpha}_\mu}{\tilde{\sigma}^2_{\alpha_\mu}}+\sum_k\sum_l\{\mu(k-1,l)-\mu_0(k-1,l)+\mu(k+1,l)-\mu_0(k+1,l)\}$$
$$\times[\mu(k,l)-\mu_0(k,l)-\beta_\mu\{\mu(k,l-1)-\mu_0(k,l-1)+\mu(k,l+1)-\mu_0(k,l+1)\}]),$$
$$(\frac{1}{\tau_\mu^2}\sum_k\sum_l\{\mu(k-1,l)-\mu_0(k-1,l)+\mu(k+1,l)-\mu_0(k+1,l)\}^2+\frac{1}{\tilde{\sigma}^2_{\alpha_\mu}})^{-1}] \tag{A. 34}$$

We use samples from $[\alpha_{\mu_{pseudo}}|\cdot]$ to obtain samples from $[\alpha_\mu|\cdot]$ in a Metropolis-Hastings sampling step within the Gibbs sampler as discussed in Section 4.3.

The full conditional distribution $[\beta_\mu|\cdot]$ is analogous to that above, as are the full conditional distri-

butions for α_a and β_a.

6.9 $[\sigma_\epsilon^2 \mid \cdot\,]$

From the discussion in Section 4.2.1 and Eqn. 40,

$$\begin{aligned}
[\sigma_\epsilon^2 \mid \cdot\,] &\propto \prod_{t=1}^{T} [\vec{\boldsymbol{Z}}_t | \mathbf{K}, \vec{\boldsymbol{Y}}_t, \sigma_\epsilon^2][\sigma_\epsilon^2 | \tilde{q}_\epsilon, \tilde{r}_\epsilon] \\
&\propto \frac{1}{(\sigma_\epsilon^2)^{\frac{mT}{2}}} \exp\{-\frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^{T} (\vec{\boldsymbol{Z}}_t - \mathbf{K}\vec{\boldsymbol{Y}}_t)'(\vec{\boldsymbol{Z}}_t - \mathbf{K}\vec{\boldsymbol{Y}}_t)\} \\
&\quad \times \frac{1}{\Gamma(\tilde{q}_\epsilon)} \tilde{r}_\epsilon^{-\tilde{q}_\epsilon} \frac{1}{(\sigma_\epsilon^2)^{\tilde{q}_\epsilon + 1}} \exp\{-\frac{1}{\tilde{r}_\epsilon \sigma_\epsilon^2}\} \\
&\propto \frac{1}{(\sigma_\epsilon^2)^{\frac{mT}{2} + \tilde{q}_\epsilon + 1}} \exp\{-\frac{1}{2\sigma_\epsilon^2} [\frac{2}{\tilde{r}_\epsilon} + \sum_{t=1}^{T} (\vec{\boldsymbol{Z}}_t - \mathbf{K}\vec{\boldsymbol{Y}}_t)'(\vec{\boldsymbol{Z}}_t - \mathbf{K}\vec{\boldsymbol{Y}}_t)]\},
\end{aligned} \tag{A. 35}$$

and,

$$\sigma_\epsilon^2 \mid \cdot \quad \sim \quad IG(\frac{mT}{2} + \tilde{q}_\epsilon, [\frac{1}{\tilde{r}_\epsilon} + \frac{1}{2} \sum_{t=1}^{T} (\vec{\boldsymbol{Z}}_t - \mathbf{K}\vec{\boldsymbol{Y}}_t)'(\vec{\boldsymbol{Z}}_t - \mathbf{K}\vec{\boldsymbol{Y}}_t)]^{-1}). \tag{A. 36}$$

6.10 $[\sigma_\gamma^2 \mid \cdot\,]$

From Eqns. 14 and 41,

$$\begin{aligned}
[\sigma_\gamma^2 \mid \cdot\,] &\propto \prod_{t=1}^{T} [\vec{\boldsymbol{Y}}_t | \vec{\boldsymbol{\mu}}, \vec{\boldsymbol{M}}_t, \vec{\boldsymbol{X}}_t, \sigma_\gamma^2][\sigma_\gamma^2 | \tilde{q}_\gamma, \tilde{r}_\gamma] \\
&\propto \frac{1}{(\sigma_\gamma^2)^{\frac{ST}{2} + \tilde{q}_\gamma + 1}} \exp\{-\frac{1}{2\sigma_\gamma^2} [\frac{2}{\tilde{r}_\gamma} + \sum_{t=1}^{T} (\vec{\boldsymbol{Y}}_t - \vec{\boldsymbol{\mu}} - \vec{\boldsymbol{M}}_t - \vec{\boldsymbol{X}}_t)' \\
&\quad \times (\vec{\boldsymbol{Y}}_t - \vec{\boldsymbol{\mu}} - \vec{\boldsymbol{M}}_t - \vec{\boldsymbol{X}}_t)]\},
\end{aligned} \tag{A. 37}$$

and

$$\sigma_\gamma^2 \mid \cdot \quad \sim \quad IG(\frac{ST}{2} + \tilde{q}_\gamma, [\frac{1}{\tilde{r}_\gamma} + \frac{1}{2}\sum_{t=1}^{T}(\vec{\boldsymbol{Y}}_t - \vec{\boldsymbol{\mu}} - \vec{\boldsymbol{M}}_t - \vec{\boldsymbol{X}}_t)' \times (\vec{\boldsymbol{Y}}_t - \vec{\boldsymbol{\mu}} - \vec{\boldsymbol{M}}_t - \vec{\boldsymbol{X}}_t)]^{-1}). \tag{A. 38}$$

6.11 $[\sigma_\eta^2 \mid \cdot\,]$

From Eqns. 20 and 40,

$$\begin{aligned}[\sigma_\eta^2 \mid \cdot\,] &\propto \prod_{t=1}^{T}[\vec{\boldsymbol{X}}_t|\mathbf{H}, \vec{\boldsymbol{X}}_{t-1}, \sigma_\eta^2][\sigma_\eta^2|\tilde{q}_\eta, \tilde{r}_\eta] \\ &\propto \frac{1}{(\sigma_\eta^2)^{\frac{ST}{2}+\tilde{q}_\eta+1}} \exp\{-\frac{1}{2\sigma_\eta^2}[\frac{2}{\tilde{r}_\eta} + \sum_{t=1}^{T}(\vec{\boldsymbol{X}}_t - \mathbf{H}\vec{\boldsymbol{X}}_{t-1})'(\vec{\boldsymbol{X}}_t - \mathbf{H}\vec{\boldsymbol{X}}_{t-1})]\},\end{aligned} \tag{A. 39}$$

and

$$\sigma_\eta^2 \mid \cdot \quad \sim \quad IG(\frac{ST}{2} + \tilde{q}_\eta, [\frac{1}{\tilde{r}_\eta} + \frac{1}{2}\sum_{t=1}^{T}(\vec{\boldsymbol{X}}_t - \mathbf{H}\vec{\boldsymbol{X}}_{t-1})'(\vec{\boldsymbol{X}}_t - \mathbf{H}\vec{\boldsymbol{X}}_{t-1})]^{-1}). \tag{A. 40}$$

6.12 $[\tau_\mu^2 \mid \cdot\,]$,$[\tau_a^2 \mid \cdot\,]$

From Eqns. 16 and 43, we obtain

$$\begin{aligned}[\tau_\mu^2 \mid \cdot\,] &\propto [\vec{\boldsymbol{\mu}}|\vec{\boldsymbol{\mu}}_0, \alpha_\mu, \beta_\mu, \tau_\mu^2][\tau_\mu^2|\tilde{q}_\mu, \tilde{r}_\mu] \\ &\propto \frac{1}{(\tau_\mu^2)^{\frac{S}{2}}} \exp\{-\frac{1}{2\tau_\mu^2}(\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)'(\mathbf{I} - \mathbf{C}_\mu)(\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)\} \\ &\quad \times \frac{1}{\Gamma(\tilde{q}_\mu)}\tilde{r}_\mu^{-\tilde{q}_\mu}\frac{1}{(\tau_\mu^2)^{\tilde{q}_\mu+1}}\exp\{-\frac{1}{\tilde{r}_\mu\tau_\mu^2}\} \\ &\propto \frac{1}{(\tau_\mu^2)^{\frac{S}{2}+\tilde{q}_\mu+1}} \exp\{-\frac{1}{2\tau_\mu^2}[\frac{2}{\tilde{r}_\mu} + (\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)'(\mathbf{I} - \mathbf{C}_\mu)(\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)]\},\end{aligned} \tag{A. 41}$$

and

$$\tau_\mu^2 \mid \cdot \quad \sim \quad IG(\frac{S}{2} + \tilde{q}_\mu, [\frac{1}{\tilde{r}_\mu} + \frac{1}{2}(\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)'(\mathbf{I} - \mathbf{C}_\mu)(\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)]^{-1}). \tag{A. 42}$$

Analogously,

$$\tau_a^2 \mid \cdot \quad \sim \quad IG(\frac{S}{2} + \tilde{q}_a, [\frac{1}{\tilde{r}_a} + \frac{1}{2}(\vec{\boldsymbol{a}} - \vec{\boldsymbol{a}}_0)'(\mathbf{I} - \mathbf{C}_a)(\vec{\boldsymbol{a}} - \vec{\boldsymbol{a}}_0)]^{-1}). \tag{A. 43}$$

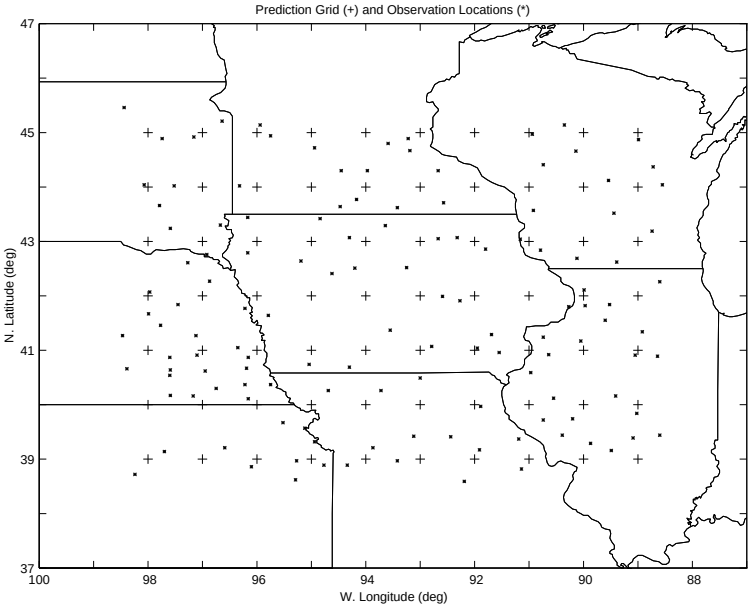

Figure 1: Prediction grid and observation locations

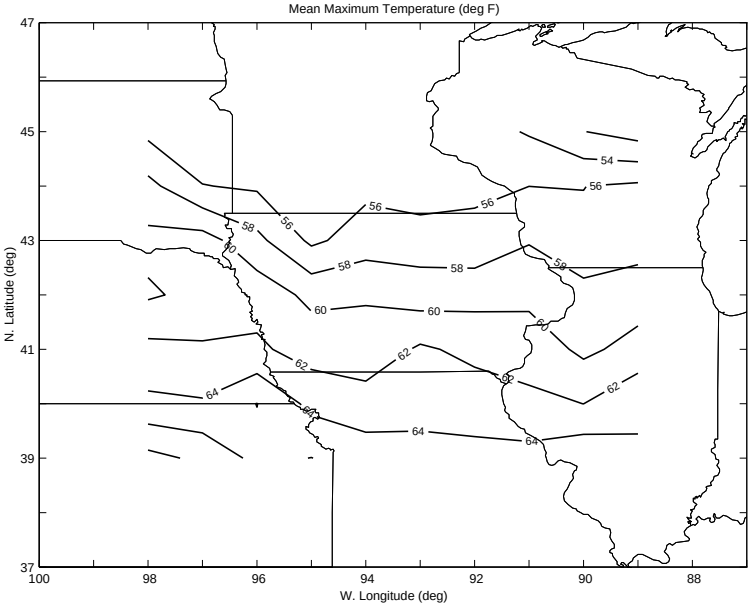

Figure 2: Contour plot of estimated means of monthly maximum temperature at observations sites (Contour interval = 2 deg F)

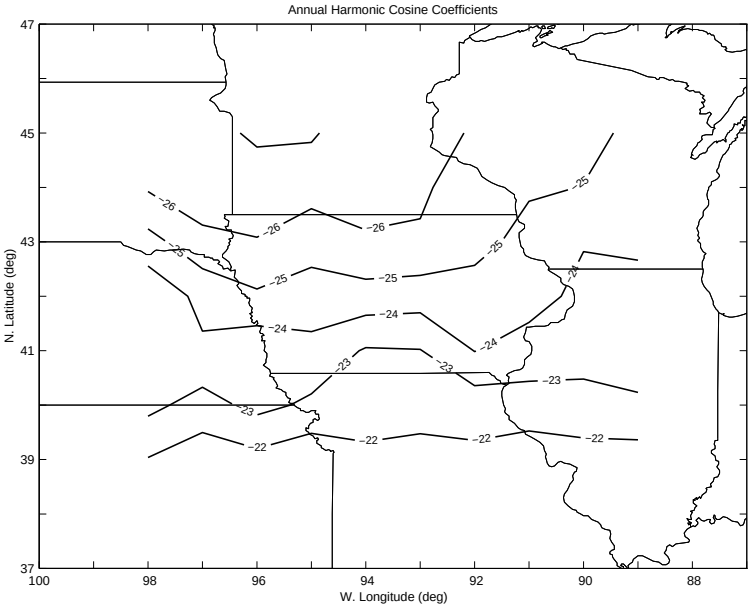

Figure 3: Contour plot of estimated cosine component of annual harmonic, estimated from deviations of site-specific temperature means (Contour interval = 1)

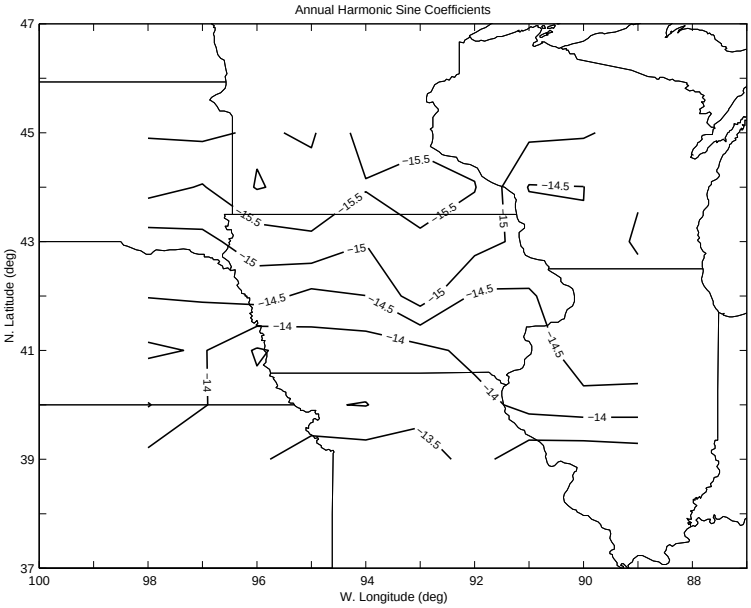

Figure 4: Contour plot of sine component of annual harmonic, estimated from deviations of site-specific temperature means (Contour interval = 0.5)

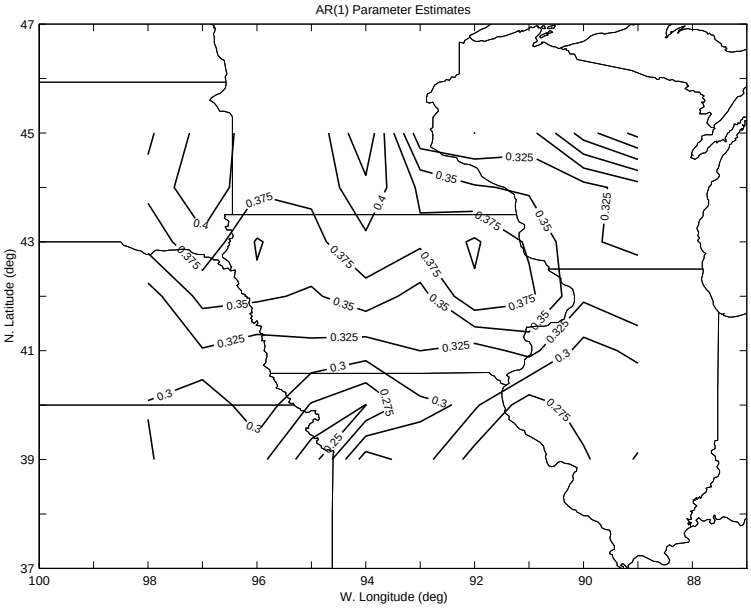

Figure 5: Contour plot of autoregressive order 1 parameters, estimated from deviations from site-specific temperature means and annual harmonics (Contour interval = 0.025)

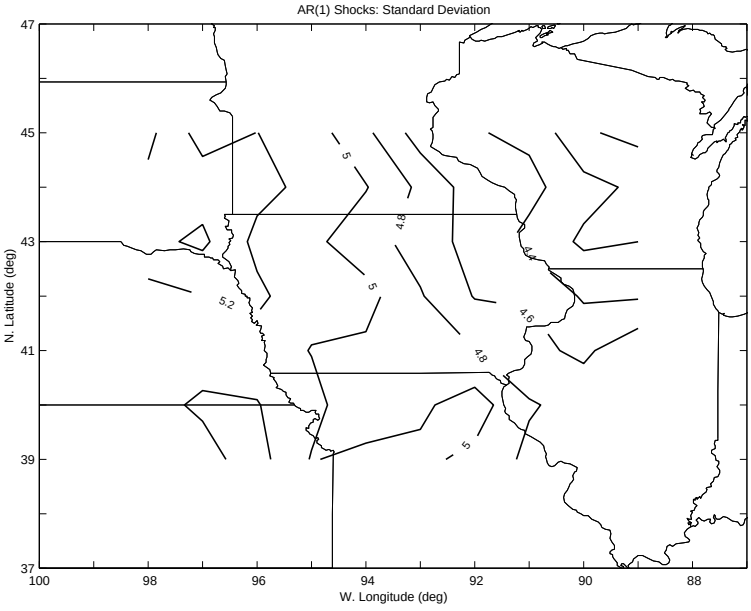

Figure 6: Contour plot of autoregressive order 1 shock standard deviations, estimated from deviations from site-specific temperature means and annual harmonics (Contour interval = 0.2 (deg F)

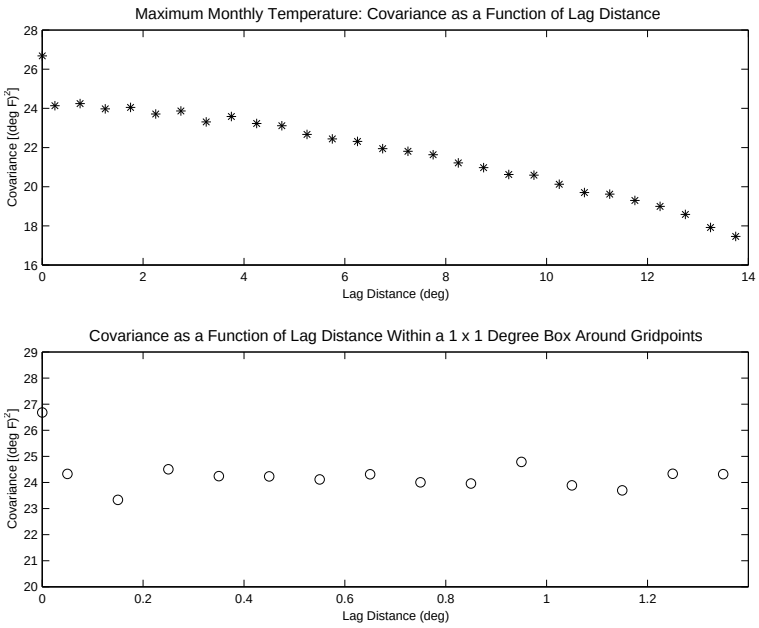

Figure 7: Maximum monthly temperature covariance estimates as a function of lag. Top panel: estimates over entire observational domain. Bottom panel: estimates within a one degree by one degree box surrounding observation locations.

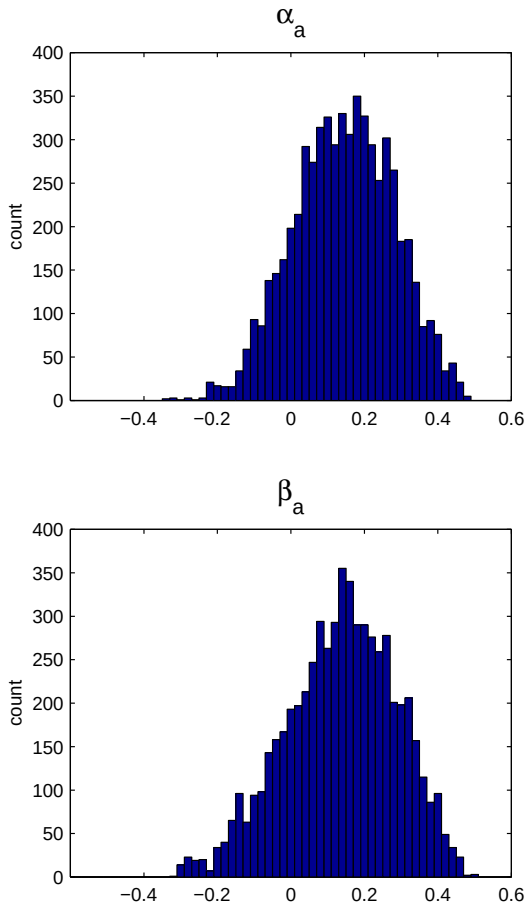

Figure 8: Histogram of Gibbs samples from the posterior distributions of α_a and β_a.

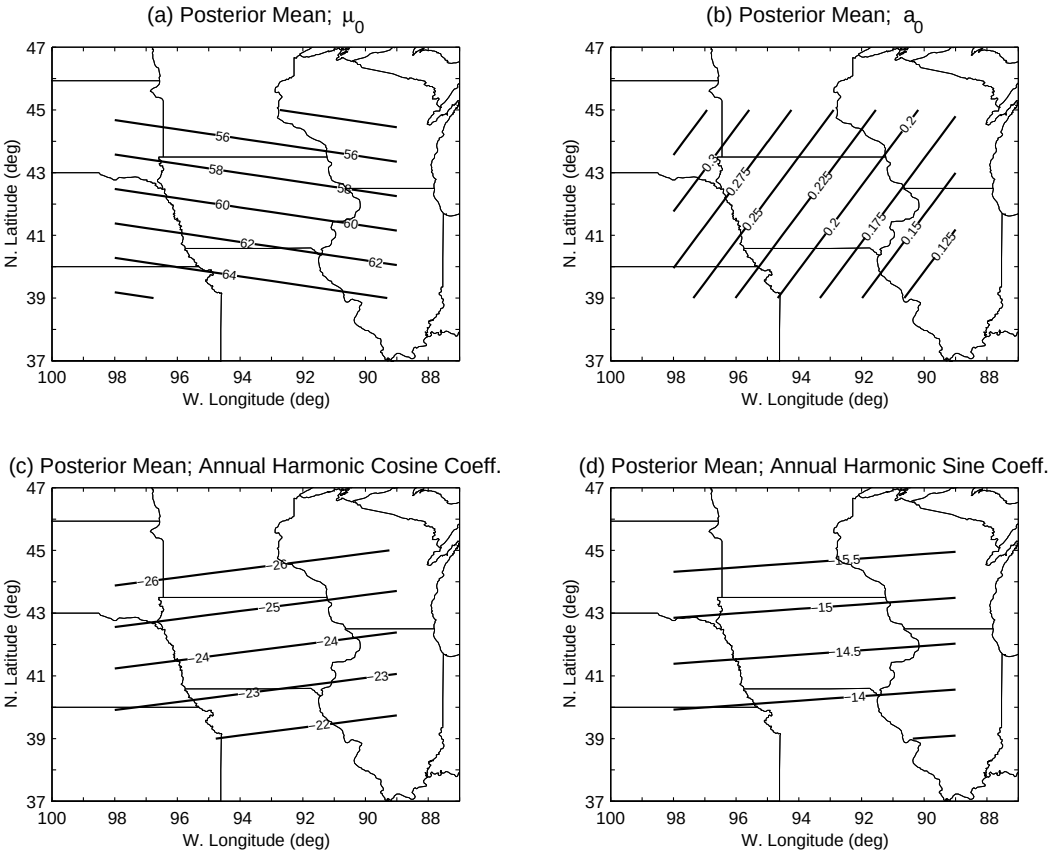

Figure 9: Linear spatial trend contour maps based on posterior means. (a) $\vec{\boldsymbol{\mu}}_0$; monthly temperature site-specific mean, MRF mean (contour interval 2 deg F). (b) $\vec{\boldsymbol{a}}_0$; same-site lagged autoregressive parameter MRF mean (contour interval .025). (c) $\vec{\boldsymbol{f}}$; annual harmonic cosine coefficients (contour interval -1). (d) $\vec{\boldsymbol{g}}$; annual harmonic sine coefficients (contour interval -.5).

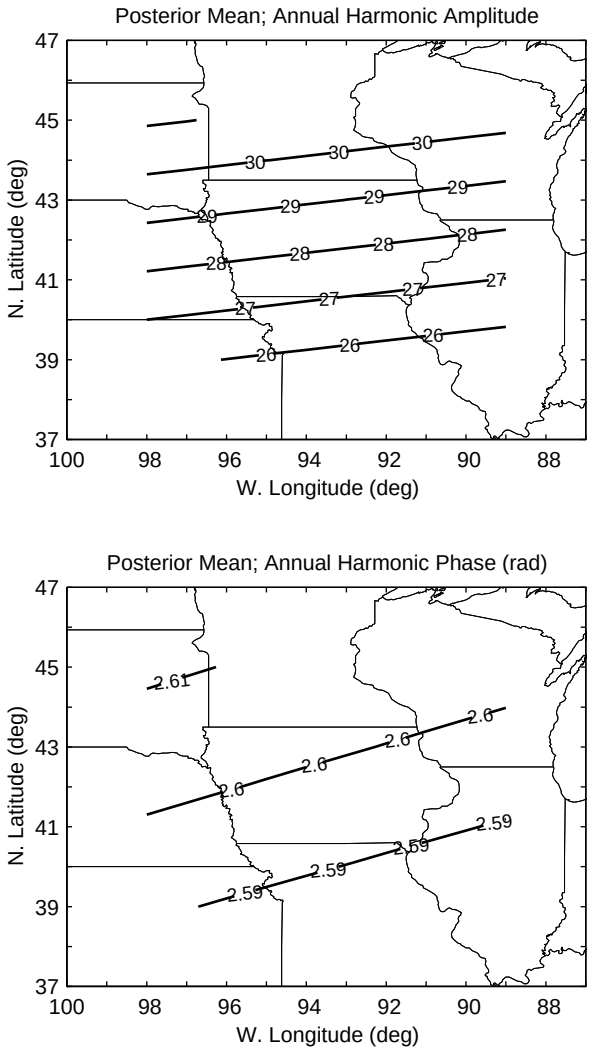

Figure 10: Linear spatial trend contour maps of annual harmonic amplitude and phase (contour intervals 1 adn 0.01, respectively).

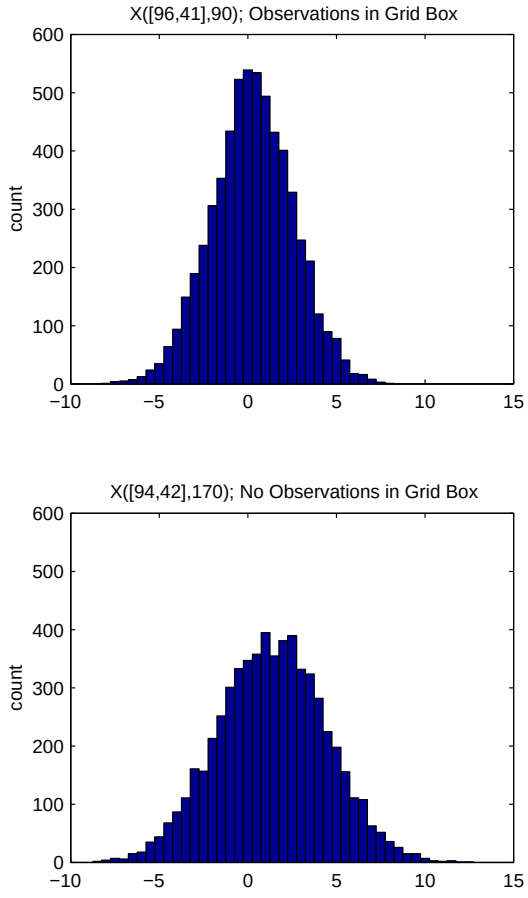

Figure 11: Histograms of Gibbs samples from prediction grid locations with observations in the grid box($X([96,41],90)$) and without observations in the grid box ($X([94,42],170)$.

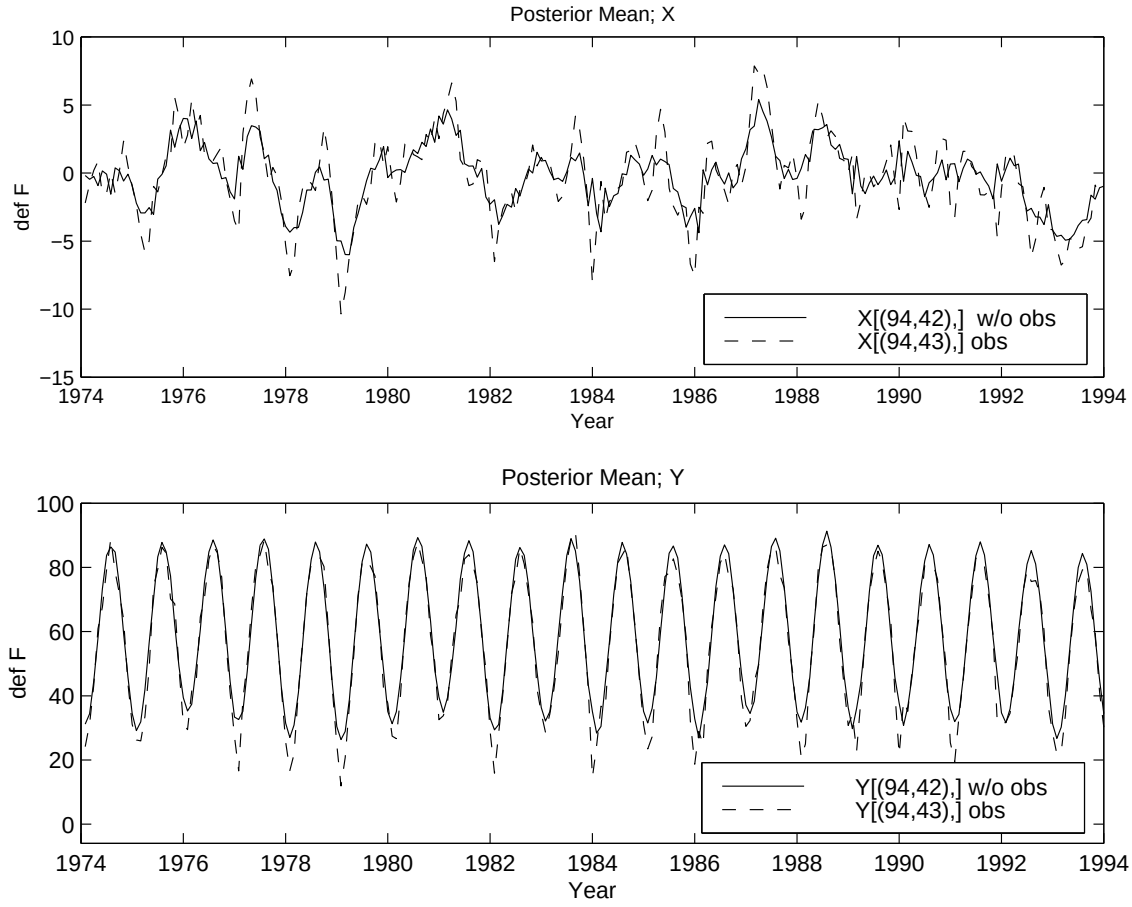

Figure 12: Posterior mean time series for X and Y (deg F). Location with observations $\{[(94,43),]\}$ and without observations $\{[(94,42),]\}$.

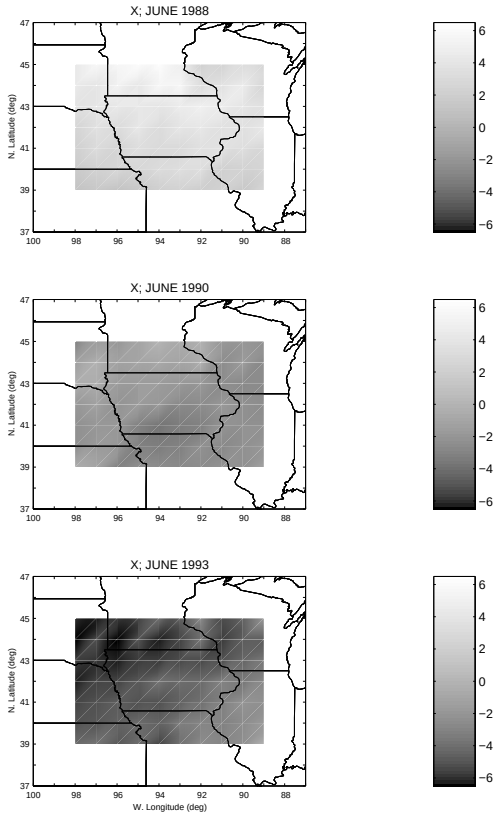

Figure 13: Posterior mean spatial maps of X process for June during 1988, 1990, and 1993 (deg F).

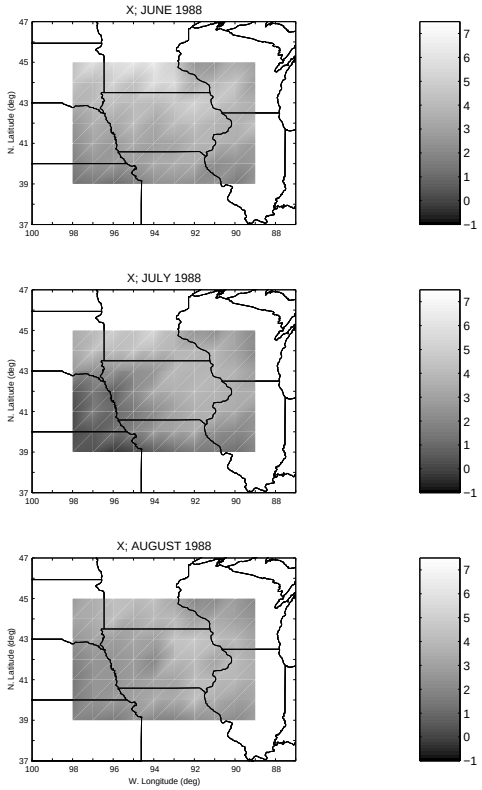

Figure 14: Posterior mean spatial maps of X process for June, July, and August during 1988 (deg F).